

THIS WEEK

EDITORIALS

MIND GAINS Brain-mapping papers add impetus to big neuroscience **p.8**

WORLD VIEW Break down borders for wildlife conservation **p.9**



GREEN CRAB Crustacean that's ready for climate change **p.11**

Brace for impacts

The latest instalment of the Fifth Assessment Report from the Intergovernmental Panel on Climate Change lays out the state of the world — and the challenges ahead.

When the Intergovernmental Panel on Climate Change (IPCC) released its previous report in 2007, some scientists and many environmentalists were still loath to talk about adapting to climate change. The policy focus was squarely on reducing greenhouse-gas emissions, and even talking about adaptation was often seen as defeatist. Thankfully, that sentiment has faded, and, although reducing emissions remains a paramount issue, climate-adaptation efforts are now under way in the private and public sectors in many countries. But as the latest instalment of the IPCC's Fifth Assessment Report — covering climate impacts, adaptation and vulnerability — makes all too clear, humanity has a long way to go in preparing for the effects that are already inevitable owing to our history, let alone for a future in which emissions continue to rise.

The report documents a range of potential impacts, from reduced agricultural yields to increased water shortages and often-unpredictable stresses on ecosystems around the globe. None of these comes as a surprise, and citizens and policy-makers would be wise to act sooner rather than later. There is no going back, regardless of the policies that might be enacted at the next United Nations climate summit in Paris in 2015, or at the meetings that follow. In that sense, this is not just a report about climate impacts. It is a comprehensive assessment of the state of the world in which we live and the direction in which it is inevitably headed. The alternatives to rapid, coordinated adaptation will be even more costly.

The latest assessment makes broad statements about the probable impacts of a warming climate on agriculture, but significant uncertainties remain at regional and global scales. Most model projections indicate reduced yields for major food crops such as rice, wheat and maize (corn) in many regions, but more than 25% of projections indicate that yields could rise by mid-century and beyond. Time does not stop in 2050, and the IPCC assigns a high confidence level to its statement that longer-term temperature increases in excess of 4°C above late-twentieth-century levels, when combined with increasing demand, would pose significant risks to food security at regional and global scales.

Scientists have their work cut out. They must continue to collect data about the ever-changing world while seeking to understand how diverse landscapes — both wild and tamed — will respond to global warming. And they must drill down to the regional level to help governments and people to make good decisions. The IPCC should consider working with local experts and bodies on detailed assessments that could then be vetted by the agency itself — a process that would engage more people and counter the impression of paternalism that sometimes shades the IPCC's work, dominated as it is by Western scientists.

In addition to filling in the knowledge gaps, scientists and the IPCC must also work to communicate clearly what is not known, be it the cause of extreme weather or the relative benefits of the suite of policy options facing governments. The current report, to the IPCC's credit, casts doubt on the models that are used to estimate economic impacts, suggesting that they are based on “disputable” assumptions. Such

projections are important when it comes to deciding who should pay which bills and when; researchers and economists must tackle these questions and more reliably determine the extent to which climate change will diminish wealth and economic growth in the long term.

“Policy-makers would be wise to act sooner rather than later.”

There is no single approach to adaptation, and strategies range from basic economic diversification to improvements in irrigation and fertilizer use. Certain regions might pursue flood protections, early-warning systems and insurance programmes, whereas others might focus on

maintenance and restoration of coastal vegetation. In the poorest countries, international financial support will be necessary. But to make it all come together, scientists and policy-makers urgently need to invest extra resources to improve environmental observations and to determine what actions make the most sense. ■

Natural decline

Few biology degrees still feature natural history. Is the naturalist a species in crisis?

What has become of the naturalists of yesteryear — the vicar with the magnifying glass and pressed flower collection, or the gentleman scientist with butterfly nets and a shotgun? Those dedicated observers of the natural world in all its complexity are still among us. But they are harder to pick out now; they are men and women, students and citizens. And they clutch not sample jars but smartphones.

In an article published late last month (J. J. Tewksbury *et al.* *BioScience* <http://doi.org/r5g>; 2014), Joshua Tewksbury, a naturalist at the University of Washington, Seattle, and 16 colleagues issue a call to arms. They chronicle the dismaying diminution of support for natural history — that branch of science that encompasses the careful observation and description of organisms and their relations to their environments. Like all good scientists, they offer the data to support their assertion.

In the United States of 1950, an undergraduate degree in biology generally required two or more courses in natural history. Today, the average number of required natural-history courses for the same degree is zero. The amount of natural-history content in biology textbooks has dropped by 40% over the past six decades. PhDs granted in natural-history-related fields are becoming ever rarer. Biological collections are on the wane as well. The number of herbaria — research collections

of plant specimens — in Europe and North America peaked in 1990.

Research in the life sciences is not created or destroyed: it simply shifts from one form to another. As natural history has been de-emphasized, molecular biology, genetics, experimental biology and ecological modelling have flourished. But here is the problem: many of those fields ultimately rely on data and specimens from natural history. Natural-history observations help to fight infectious diseases that cycle through different species, to identify promising leads for drug discovery, to manage fisheries and forests and other natural resources and to conserve species and ecosystems.

As Tewksbury and his colleagues write: “Direct knowledge of organisms — what they are, where they live, what they eat, why they behave the way they do, how they die — remains vital to science and society.” The best algorithms in the world will fail to guide our action accurately if they are not based on a firm understanding of what is out there and what it’s up to.

Revitalizing natural history will require tweaking the research incentives of grants and academic tenure. The *BioScience* article is right to call for natural historians to go out and stress the enduring importance of their craft to universities, funding agencies, foundations and the public. No biology student should get a diploma without at least a single course in identifying organisms and learning basic techniques for observing and recording data about them. Top journals should publish excellent natural history; the revived ‘Natural History Miscellany’ section in *American Naturalist* is a good first step.

Natural history itself can adapt to help. It should continue to expand beyond the elite, lone naturalist. New digital tools, including mobile

versions of field guides (such as the Leafsnap app, which can identify tree species from photographs, and the Chirp! app, which helps users to recognize bird songs), are lowering the bar for entry for those without training. And digital data repositories — such as eBird,

“No biology student should get a diploma without at least a single course in identifying organisms.”

created by the Cornell Lab of Ornithology in Ithaca, New York, and the New York-based National Audubon Society — mean that today’s naturalists can share and compare their observations. These tools can be used by the general public to build big data sets, which can feed into experiments or models.

As scientists from Yale University point out in a Comment piece on page 33, such data sets are also crucial for other purposes: to hold to account the official government figures that, for one reason or another, do not accurately reflect the situation on the ground, in the air or in the seas.

Natural history has never been just about the science. It is a craft and a passion with its own immediate aesthetic and visceral pleasures, the epitome of a positive relationship with nature. The smartphone, as the most ubiquitous representative of an increasingly digital culture, has often been held up as the pernicious opposite of a direct relationship with the natural world. But technology can be used as a tool to draw us closer to nature as well as a screen to block our view.

The dedicated observers are still there. They tramp through the woods on cold winter nights, their breath visible in the moonlight. They play the calls of the great horned owl on their smartphones. And the great horned owls call back. ■

Brain waves

Above the ‘big neuroscience’ commotion, literature plays its part.

“Literature was not born the day when a boy crying ‘wolf, wolf’ came running out of the Neanderthal valley with a big grey wolf at his heels,” wrote novelist Vladimir Nabokov. Instead, he argued, it was born “on the day when a boy came crying ‘wolf, wolf’ and there was no wolf behind him”.

The French consciousness-research pioneer Stanislaw Dehaene uses this quote in his new book, *Consciousness and the Brain*, in which he describes his ‘global neuronal workspace’ theory, elaborated together with Jean-Pierre Changeux through modelling a 20-year series of daring experiments probing conscious and unconscious perception in humans. Only since brain imaging and other tools have allowed us to view the human brain at work has it become ‘respectable’ to try to pin down consciousness, and to debate how the human mind has allowed the development of intellectual pursuits as sophisticated as literature.

The Dehaene–Changeux theory holds that awareness moves from subconscious to conscious only when we pay attention to specific sets of information in our brains: images, memory, emotional state. These briefly come together in a limited-capacity workspace, ready to broadcast to all brain regions through axons. This theoretical workspace is where consciousness emerges; where, for example, a storyteller may invent a fictitious scene of deceit, such as the boy who cried wolf.

Dehaene quotes Nabokov often in his books, with good reason. The poetic, multilingual novelist and entomologist often pondered eloquently on the state of being conscious. Understanding consciousness and the mind may take a century, but it stands as an irresistible beacon. Other goals, such as understanding, fixing or ameliorating neurodegenerative or psychiatric diseases, may be ‘only’ decades away.

Many regret, but few doubt, that the long haul towards these goals requires a cultural shift in neuroscience research, from small to big

science. Indeed, ‘big neuroscience’ has already begun. Last year, Europe formally launched its highly ambitious Human Brain Project (HBP), which aims to simulate the human brain in a supercomputer. It already has 32 partners across 13 countries. And the Brain Research through Advancing Innovative Neurotechnologies (BRAIN) Initiative, announced by US President Barack Obama on 2 April last year, will soon begin distributing money.

The privately funded Allen Institute for Brain Science in Seattle, Washington, has been doing big neuroscience for more than a decade, producing systematic anatomical and brain maps, mostly in the mouse, and is now starting to map functions of the component neurons. All of its maps are publicly available. It is a happy coincidence that *Nature* has published two brain-mapping papers from Allen scientists on the anniversary of Obama’s announcement (see <http://dx.doi.org/10.1038/nature13185>; <http://dx.doi.org/10.1038/nature13186>; 2014). One of the maps is the first gene-expression atlas of the entire developing human brain. The other is a mouse ‘connectome’ — the first brain-wide neuronal-connectivity map for a mammalian species — that will guide the initial modelling of the HBP, which is beginning with the mouse brain.

But the path to generating and modelling the data needed to crack the codes of the brain will not be smooth. Already, the HBP has annoyed researchers by not funding the generation of data in non-human primates in its first phase, perhaps fearing a political backlash. But monkey data will be needed as a bridge between the mouse and human brain.

As big neuroscience advances, the Dehaene–Changeux theory may be proved wrong. So, too, may the more abstractly mathematical ‘integrated information theory’ of consciousness preferred by the Allen Institute’s chief scientific officer Christof Koch. That is the process of science. Koch complains in his review of Dehaene’s new book in *Science* that the workspace theory limits itself too much to the waves of electrical activity in the brain that experiments pick up, and fails to explain the ‘why’ of consciousness (see <http://doi.org/r5q>; 2014).

Koch counter-quotes Nabokov: “The breaking of a wave cannot explain the whole sea.”

As big-brain programmes navigate their thorny early years, it is good to be so neatly reminded of their ultimate goal. ■

➔ **NATURE.COM**

To comment online,
click on Editorials at:
go.nature.com/xhunqv

CLARISSE M. HART/HARVARD FOREST



Political borders should not hamper wildlife

Given the lack of global legislation, nations should work hard to establish cross-border protections for vulnerable species, says Aaron M. Ellison.

The Republic of Crimea looks set to become the latest new nation state to emerge from ethnic division and geopolitical gamesmanship. Political action and human bloodshed may subside after the dust has settled, but many constituents will lack representation. Those are the millions of non-human species that occupy the fragmenting territories. Who speaks for them, and what protections do they have at existing borders or when borders change?

Put simply, new, unstable countries do not protect biodiversity and habitat as well as nations that have strong governance structures. Only laws and statutes protect biodiversity, and these differ on either side of nearly every border.

Ukraine has been part of the Pan-European Biological and Landscape Diversity Strategy since the programme's inception, but the Russian Federation's participation has varied widely. There (and in the Commonwealth of Independent States), protection of biodiversity is overseen by the Russian Ministry of Natural Resources and Environment, and relies on listings in the *Red Data Book of the Russian Federation*, which is nearly two decades old. So, as Crimea severs its links with Ukraine and forges new ones with Russia, what will happen to the endangered plants and animals in the Kara Dag Nature Reserve on the Crimean coast, or the endemic species of the Askania-Nova Biosphere Reserve on the border between Ukraine and Crimea?

There are many examples of steep declines in species populations after political division. The range and population size of the endangered Kashmir markhor (*Capra falconeri falconeri*) in India have contracted by 60% since the 1947 partition that formed Pakistan — not least because of the 500-kilometre fence that India erected to mark the de facto border.

In other cases it is hard even to keep track. When Yugoslavia dissolved in 1992 into Serbia, Montenegro, Slovenia, Macedonia, Croatia and Bosnia-Herzegovina, planning for the conservation and management of Yugoslavia's fragmented population of regionally threatened or endangered brown bears (*Ursus arctos arctos*) was taken over by agencies in six countries, each of which had different species-protection acts. Borders, language barriers and political constraints mean that data on population size and species status are now collected only by country, not across the entire range of the bear. Such fragmented data will not help those charged with ensuring the survival of the species. And without reliable data, conservation is likely to continue to be cast as the enemy of economic development.

Of course, these same issues arise across borders that have been stable for some time. When a grizzly bear (*Ursus arctos horribilis*) in Waterton Glacier International Peace Park wanders from

Montana to Alberta, it loses its protection under the US Endangered Species Act. The United States considers the grizzly to be regionally threatened, but Canada lists it only as a species of special concern.

Such cross-border transits are becoming increasingly difficult: the proliferation of fences between countries not only prevents illegal immigration but also slices in two the habitats of many species and prevents migrations. Populations of the globally endangered Arroyo toad (*Anaxyrus californicus*) and black-spotted newt (*Notophthalmus meridionalis*), for example, have been split into potentially unsustainable subpopulations by the fence between the United States and Mexico. Similarly, Arabian gazelles (*Gazella arabica*), striped hyenas (*Hyaena hyaena*) and sand cats (*Felis margarita*) find their historic ranges and migration routes blocked by border walls in the Middle East.

Civil wars (and the creation of new nation states) have become much more frequent since the Second World War, and they not only cause immediate environmental damage, but also lead to long-lasting differences in cross-border conservation programmes. Peaceful transitions may minimize the immediate impacts, but challenges will remain in the long term. In the United Kingdom, for example, many rare or endangered butterflies, birds and plants are moving from England into Scotland as the climate warms. If Scotland votes for independence, will these species retain their current levels of protection?

Conservation biologists have repeatedly called for global, or at least cross-border, systems of protected areas. Long-standing international treaties, such as the Convention on Biological

Diversity (CBD), make similar pronouncements. But although the CBD has guidelines and suggestions for the conservation of biological diversity, it provides no legal protection for threatened or endangered species. And even if the CBD could be given regulatory power, every time a new country is formed, treaties have to be reopened, renegotiated and re-ratified; history suggests that the result tends to be new treaties or laws with more exceptions and weaker protections.

Still, the concept of transboundary protected areas, or Peace Parks, has strong appeal. Peace parks, such as the La Amistad International Park established on the border between Costa Rica and Panama in 1988, not only protect biodiversity but also help to defuse border tensions and conflicts. Such reserves can be set up bilaterally, without complex multinational agreements. If Crimea and Ukraine do end up parting ways, then they must set up a similar cross-boundary haven for wildlife. ■

Aaron M. Ellison is a senior research fellow at Harvard University in Cambridge, Massachusetts.
e-mail: aellison@fas.harvard.edu

WHAT PROTECTIONS DO SPECIES HAVE AT EXISTING BORDERS OR WHEN BORDERS CHANGE?

➔ NATURE.COM
Discuss this article
online at:
go.nature.com/67zvvg

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

NEUROBIOLOGY

Brain map reveals behaviour links

An “atlas” of the fruit-fly brain is the largest yet to map regions that encode specific behaviours, such as walking backwards.

Carey Priebe of Johns Hopkins University in Baltimore, Maryland, and Marta Zlatić of the Howard Hughes Medical Institute's Janelia Farm Research Campus in Ashburn, Virginia, and their colleagues engineered fruit-fly larvae so that the insects' neurons fired when hit with a beam of light. The researchers stimulated more than 1,000 different neuronal pathways in nearly 38,000 *Drosophila* flies, and recorded how the flies responded.

They were able to determine 29 different behaviours, such as turning to avoid an obstacle, and mapped which neurons seemed to control each behaviour.

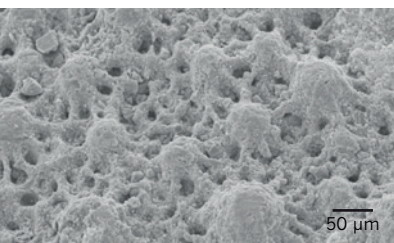
Science <http://doi.org/r4t> (2014)

PALAEONTOLOGY

Ancient starfish spotted predators

Sea stars and some other echinoderms might have had complex visual systems for roughly the past 80 million years.

Some existing echinoderms, such as brittle stars, are covered in crystal calcite microlenses that are sensitive to light. To



determine the evolutionary history of these structures, Przemysław Gorzelak at the Polish Academy of Sciences in Warsaw and his team analysed 75-million-year-old brittle-star and starfish fossils using a scanning electron microscope. Both kinds of fossil contained structures (pictured) that matched modern echinoderms' microlenses in size and shape.

After an explosion in the diversity of fish and crustacean predators began around 80 million years ago, echinoderms may have

developed visual systems to avoid such predators, the researchers say.

Nature Commun. 5, 3576 (2014)

NEUROSCIENCE

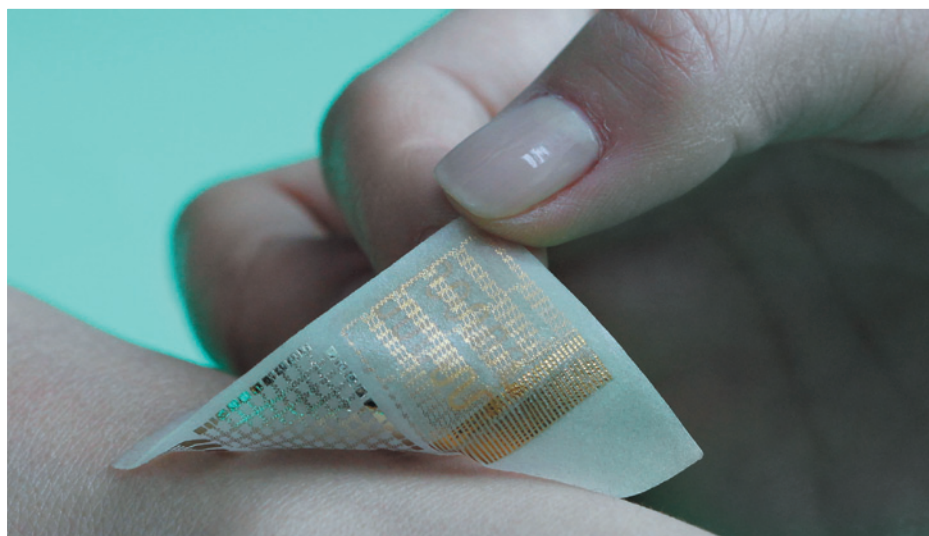
Why babies live hand to mouth

The reason that newborns put their hands into their mouths is probably because this action is hard-wired into the brain as a basic unit of movement.

Angela Sirigu at the French National Centre for Scientific Research in Bron and her

colleagues used electrodes to stimulate nearly 150 sites in the precentral gyrus — the brain region that controls voluntary movement — in 26 people undergoing brain surgery. When the researchers stimulated ten specific sites in nine of the participants, including two three-year-olds, the volunteers each moved their closing hand towards their opening mouth.

The authors speculate that the hard-wiring of these movements in the brain means that they are an evolutionarily important behaviour that



ELECTRONICS

Stick-on skin sensor measures motion

A wearable device as thin as a temporary tattoo can measure, store and transmit data on muscle activity, and release embedded drugs into the wearer's skin.

Dae-Hyeong Kim at Seoul National University in South Korea and his colleagues built their device (pictured) by placing stretchable layers of nanomaterials onto an elastomeric polymer material designed to mimic the softness and flexibility of skin. The nanomaterials acted as strain and temperature sensors, memory modules, microheaters

and drug carriers. The authors showed that when the device was applied to human skin, it remained in place and deformed with the skin. It measured simulated hand tremors, and delivered drugs through the skin when the tiny heaters generated enough heat.

The device used a wired connection for power supply and data transfer, but the team aims to develop a wireless version for use by patients with movement disorders.

Nature Nanotechnol. <http://dx.doi.org/10.1038/nnano.2014.38> (2014)

DONGHEE SON/JONGHA LEE

PRZEMYSŁAW GORZELAK/
POLISH ACADEMY OF SCIENCES

allows babies to put things in their mouth accurately, even at a time when they have generally poor motor control. *Proc. Natl Acad. Sci. USA* <http://dx.doi.org/10.1073/pnas.1321909111> (2014)

ANIMAL BEHAVIOUR

Whale dives into record books

A Cuvier's beaked whale has set a new record for the deepest known dive by a mammal, reaching 2,992 metres.

Gregory Schorr at the Cascadia Research Collective in Olympia, Washington, and his colleagues attached satellite tags to eight Cuvier's beaked whales (*Ziphius cavirostris*) and collected more than 3,700 hours of data on the whales' movements off the California coast. One individual was in a dive for 137.5 minutes, nearly 18 minutes longer than the previous deep-diving record-holder, the southern elephant seal (*Mirounga leonina*). Another whale beat the seal's record depth by more than 600 metres.

This behaviour could be atypical, and might be a result of heavy naval sonar use in part of the study region, the authors say.

PLoS ONE 9, e92633 (2014)

CANCER GENETICS

Cancer survives by silencing a gene

Breast cancer can become resistant to treatment by co-opting a gene-silencing mechanism, reports a team led by Steffi Oesterreich at the University of Pittsburgh in Pennsylvania.

The hormone oestrogen, which drives many breast cancers, dampens the activity of the tumour-fighting gene *HOXC10*, and drugs called aromatase inhibitors free the gene from this repression. But in a genome-wide screen of human breast-cancer cells, the team found that these drugs can also lead to a type of

epigenetic modification called methylation — the addition of methyl groups to DNA without changing its sequence — across the genome. This ultimately silences *HOXC10*, rendering breast-cancer cells resistant to aromatase inhibitors.

Blocking the methylation activity associated with aromatase-inhibitor treatment might delay or prevent resistance to therapy, the authors say.

Sci. Transl. Med. 6, 229ra41 (2014)

NEUROSCIENCE

A broken channel in Huntington's

In Huntington's disease, neurons become more excitable and die. Researchers have now found a faulty ion channel in astrocytes, another type of brain cell, that could be contributing to this.

Baljit Khakh and Michael Sofroniew at the University of California, Los Angeles, and their colleagues studied two mouse models of Huntington's disease, and show that channels that allow potassium ions to cross the cell membrane were dysfunctional in astrocytes containing the Huntington's disease proteins. This caused potassium levels outside the cells to rise, making the membranes of nearby neurons more excitable.

The authors reversed these defects by replacing the faulty channels with functional ones. In one of the mouse models, channel replacement allowed the animals to walk more normally and live longer than mice with defective channels. *Nature Neurosci.* <http://dx.doi.org/10.1038/nn.3691> (2014)

AGRICULTURE

Cattle tamed by moving and mixing

Cattle were domesticated in the Middle East before being brought to Africa by migrating humans some 10,000 years ago.

Researchers previously thought that African cattle were domesticated there. To

COMMUNITY CHOICE

The most viewed papers in science

ECOLOGY

City birds and plants in decline

HIGHLY READ
on *Proc. R. Soc. B*
online in February

Most bird and plant species in cities are native to those areas, but their numbers are rapidly decreasing around the world.

Myla Aronson at Rutgers University in New Brunswick, New Jersey, and her colleagues compiled and analysed data for birds in 54 cities and for plants in 110 cities, mainly in North America and Europe — the largest collection of urban biodiversity data so far.

The authors found that cities support just 8% of bird species and 25% of plant species that are found in non-urban areas. Human-related factors, such as land use and city age, seem to have a greater effect on bird and plant populations than do natural factors such as climate and geography.

Urban planning that emphasizes native habitats could better support biodiversity, the authors say.

Proc. R. Soc. B 281, 20133330 (2014)

better understand this history, Jared Decker and Jeremy Taylor at the University of Missouri in Columbia and their colleagues analysed the DNA of 134 breeds of domesticated cattle to establish the relationships between them. The authors found that cattle imported from the Middle East bred with wild species to produce the African animals seen today.

Mixing of native cattle with imported breeds occurred extensively worldwide. For example, American feral breeds are descended from cattle that were brought in from Spain and India.

PLoS Genet. 10, e1004254 (2014)

tolerance of crabs by measuring the cardiac function of animals at seven sites in Europe and eastern North America with widely varying temperatures. They found that the crabs could withstand warmer waters before their cardiac function was compromised, compared to native crustaceans in many places. Crabs also thrived in colder habitats, and acclimatized quickly to temperature shifts.

Temperature tolerance has probably allowed this crab to spread rapidly along the coasts of North America, say the authors.

J. Exp. Biol. 217, 1129–1138 (2014)

ECOLOGY

Crabs ready for climate change

Cold-blooded animals are especially at risk as the climate warms, but the remarkable temperature tolerance of the European green crab (*Carcinus maenas*; pictured) bodes well for the animal.

Carolyn Tepolt and George Somero at Stanford University's Hopkins Marine Station in Pacific Grove, California, studied the temperature



NATURE.COM

For the latest research published by Nature visit:

www.nature.com/latestresearch

SEVEN DAYS

The news in brief

EVENTS

Telescope progress

The first of 64 huge antennas that will eventually make up the world's biggest radio telescope opened for business in South Africa on 27 March. The full array of the 19.5-metre antennas, which will make up the MeerKAT telescope under construction in the Karoo region, will be operational by 2017. The project is the first stage of the much larger planned Square Kilometer Array (SKA) of radio telescopes that will be located in South Africa and Australia. The first stage of SKA's Australian component will add 60 dishes to an existing 36-dish array in Western Australia.

Ebola spreads

An outbreak of the Ebola virus that began in southern Guinea has reached the capital, Conakry, and has moved into neighbouring Liberia. Guinean health officials have reported 78 deaths out of 122 suspected or confirmed cases, the World Health Organization (WHO) said on 30 March. Four people have died in Liberia among seven possible and confirmed cases. All of the Liberian cases were contracted in Guinea, said the WHO.

NUMBER CRUNCH

7 million

The number of people who died in 2012 as a result of exposure to air pollution, according to a report released by the World Health Organization on 25 March. The figure more than doubles previous estimates. Air pollution is now the largest single environmental health risk, the agency says.



TIM WATERS/SEA SHEPHERD AUSTRALIA/NEWSCOM

Court halts Japan's whale hunt

Japan's Antarctic whale hunt is not scientific and so must be halted, ruled the International Court of Justice on 31 March in The Hague in the Netherlands. Japan has maintained that it catches whales for research — which is allowed under international law — but opponents say that the actions amount to commercial whaling, which is effectively banned. The court found that the research programme

has produced limited scientific findings, and that Japan provided a weak explanation for the number of animals it killed. It is unclear whether Japan will now cease its hunts. Whaling elsewhere continues, for example in Norway and Iceland, which oppose the ban and do not claim a scientific justification for their whale hunts. See go.nature.com/rlu3iz for more.

POLICY

Cut threat to health

The United States' contribution to global health research and development is being undermined by funding restrictions to the US National Institutes of Health (NIH). The warning comes in a 27 March report from the Washington DC-based Global Health Technologies Coalition, which represents more than 25 non-profit organizations. NIH funding cuts have forced the closure of clinical trials examining a new tuberculosis drug and the scaling back of programmes providing poor countries with treatments for HIV/AIDS and tools to control malaria.

Water protection

Wetlands and streams in the United States are set to be given greater protection under the Clean Water Act. On 25 March, the Environmental Protection Agency and the US Army Corps of Engineers sought to clarify the scope of the act, proposing that wetlands and streams specifically should be protected. Environmental groups have been calling for such clarification for a decade. The proposal is open for public comment until the end of June.

Methane plan

The US government published on 28 March a national strategy to cut methane

emissions, as part of the Climate Action Plan that President Barack Obama announced last June. The strategy includes proposals for updated standards, set to be issued this summer, to cut emissions of the potent greenhouse gas from landfills. The government also announced a forthcoming Biogas Roadmap that will outline voluntary strategies to reduce methane emissions from the dairy industry by 25% by 2020.

Warming world

Global warming is already affecting ecosystems across Earth, from freshwater systems to urban environments, and its effects are poised to increase in

NASA/JAXA the coming decades, said the Intergovernmental Panel on Climate Change on 31 March. The report was the second instalment of the panel's fifth climate assessment, focusing on global warming's impacts and the need for humans to adapt to a changing climate. The last instalment of the report is scheduled for release on 13 April. See page 7 for more.

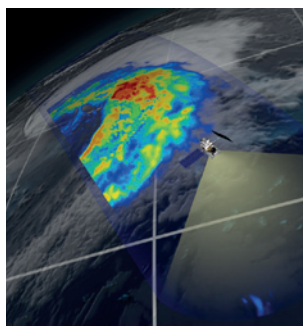
Acidic waters

The US government should establish a national programme to coordinate research on ocean acidification, a federal advisory group said on 26 March. Developing technologies to monitor the pH of the ocean and how it affects marine food webs are among the goals that the programme should pursue, said the Interagency Working Group on Ocean Acidification. The group warned that the chemistry of the world's oceans is changing faster now than at any other time in the past 20 million years.

RESEARCH

Cyclone captured

Bands of heavy rain stand out in red in this image of a cyclone in the northwest Pacific Ocean on 10 March, one of the first glimpses from the US–Japanese Global Precipitation Measurement



(GPM) Core Observatory satellite. GPM officials released the images on 25 March, less than four weeks after the mission launched. The cyclone pictured shows off the GPM's ability to measure a variety of precipitation in a single storm, including snowflakes. Rain is shown in red, yellow and light blue, and snow in dark blue (upper left).

Drug discovery

Two planned UK research centres will bring together academia and industry to discover medicines. The Centre for Therapeutic Target Validation, announced on 27 March, will emerge from a collaboration between drug company GlaxoSmithKline, headquartered in London, the Wellcome Trust Sanger Institute and the European Bioinformatics Institute, both in Hinxton, UK. On 31 March, the UK Medical Research Council announced

a separate collaboration with London-based drug company AstraZeneca to open the Centre for Lead Discovery in 2016, aimed at better understanding diseases.

Misconduct finding

A committee investigating problems with two high-profile stem-cell papers says that lead researcher Haruko Obokata manipulated data in an intentionally misleading fashion. Obokata, of the RIKEN Center for Developmental Biology in Kobe, Japan, is guilty of scientific misconduct, said the RIKEN-scientist-led committee at a press conference on 1 April. Obokata was not present, but said in a written statement that she planned to appeal. The papers, which claimed that a simple method reprogrammed mature cells into an embryonic state, were published in *Nature* on 30 January. See go.nature.com/cltkcg for more.

Site selection

European scientists have picked four potential landing sites for a rover designed to search for life on Mars. The ExoMars rover, part of a joint mission of the European Space Agency (ESA) and Russia's space agency, Roscosmos, is scheduled to land on the red planet in early 2019. At

COMING UP

3 APRIL

The Sentinel-1A satellite launches. It is the first of two satellites in the initial phase of Europe's Copernicus global-monitoring programme, and carries an advanced radar to provide all-weather, 24-hour images of Earth.

go.nature.com/ijm6cd

5–9 APRIL

The American Association for Cancer Research meets in San Diego, California, to discuss and celebrate progress in turning basic-science discoveries into treatments.

go.nature.com/c1mtj3

a workshop on 27 March at ESA's European Space Astronomy Centre near Madrid, four sites — Hypanis Vallis, Oxia Planum, Mawrth Vallis and Oxia Palus — emerged as favourites from eight proposed sites. An expert group will now consider the proposals before announcing a formal shortlist in June. See page 19 for more.

AWARDS

Maths prize

The Norwegian Academy of Science and Letters has awarded the 2014 Abel Prize to Yakov Sinai, a Russian-born mathematical physicist at Princeton University in New Jersey. Sinai is most famous for his work in the field of complex dynamical systems, which includes chaos theory. The Abel Prize is one of the most prestigious prizes in mathematics and comes with a monetary award of 6 million Norwegian kroner (US\$1 million). See go.nature.com/ttwwwc for more.

► NATURE.COM

For daily news updates see:

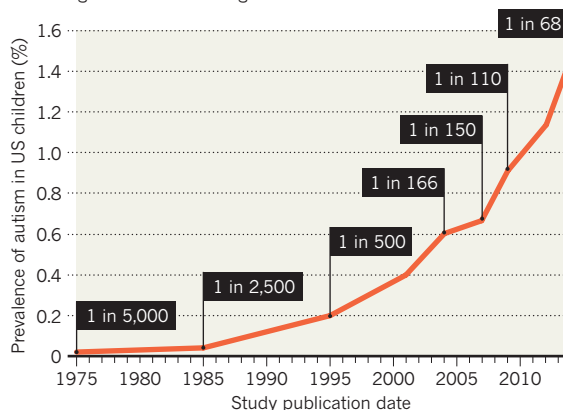
www.nature.com/news

TREND WATCH

Autism diagnoses rose to 1 in 68 US children (1.2 million in total) in 2010, up from 1 in 110 in 2006, according to a report released on 27 March by the Centers for Disease Control and Prevention in Atlanta, Georgia. Nearly half of the children who are diagnosed with autism have average or above average intelligence, up from one-third in 2002. It is unclear whether more-intelligent children are developing the condition, or whether they are being diagnosed more often than in the past (see *Nature* 479, 22–24; 2011).

AUTISM DIAGNOSES RISING

Almost 1.5% of US children are now diagnosed with autism, according to data from 11 regions in the United States.



NEWS IN FOCUS

SPACE Landing sites chosen for European mission to find life on Mars **p.19**

CLIMATE Clouds loom over El Niño weather forecasting systems **p.20**

EPIDEMIOLOGY Epigenetic studies offer fresh insight into disease **p.22**

COSMOLOGY John Kovac spies on the birth of the Universe **p.28**



ALEXEI PAVLISHAK/ITAR-TASS



The Crimean Astrophysical Observatory in Nauchny is one of many institutes in Crimea that will now be taken over by the Russian Academy of Sciences.

POLICY

Ukraine's science in turmoil

Crimean institutions put their future in Russia's hands as Ukraine attempts research reforms.

BY ALISON ABBOTT

With political chaos playing out around them, researchers in Ukraine and the freshly annexed region of Crimea are facing their own upheavals. Many Crimean institutes are looking to their new Russian master for signs of what is to come, and Ukraine's scientists can already see cuts heading their way.

The interim government in Ukraine, set up after the ousting of President Viktor Yanukovich on 22 February, has now declared a 100-day programme to reform education and research along meritocratic Western lines before the presidential elections on 25 May. But the country's austerity budget, agreed on 27 March, will see researchers lose up to one-third of their already low funding (which totalled US\$227 million in 2013) as Ukraine seeks to secure foreign loans to prioritize support for the army.

"Things are not going well," says Nataliya

Shulga, a molecular biologist at the National Pedagogical Dragomanov University in Kiev, who runs the Ukrainian Science Club, a lobby group.

Unsure of what the future holds, research institutes and universities in Crimea have requested takeover by Russian institutions, the relative wealth and stability of which are attractive. And Ukraine's interim government now accepts that its reform programme will not cover Crimean scientists. The Crimean Astrophysical Observatory in Nauchny, part of the University of Kiev, is one of around 100 science centres that are transferring to the Russian Academy of Sciences. Observatory staff received their final Ukrainian pay cheques last month.

For some, this is unsettling. "We really don't know what will happen at all, and we are reluctant to orientate ourselves to internal grants in Russia because of the international sanctions," says zoologist Gregory Prokopov from

Taurida National V. I. Vernadsky University in Simferopol, Crimea.

But many Crimean scientists are more than happy about the change. Alla Rostopchina, director of the Crimean Astrophysical Observatory, says that it was an immediate decision to request admission to the Russian Academy of Sciences. The observatory, she says, has "long and strong" relationships with Russian astronomy, and she anticipates practical benefits. "First of all we hope to update our instrumentation, and we look forward to fuller participation in Russian space projects," she says. But she also hopes to maintain collaborations with Western scientists.

Yet the shift eastwards might not be straightforward. At a meeting of the Russian Academy of Sciences' executive committee on 25 March, its president Vladimir Fortov pledged not to interfere with the academic work of its new Crimean institutes, but noted that it would cost the academy 5 billion rubles ►

► (US\$140 million) a year to operate them. And Yaroslav Yatskiv, director of the Main Astronomical Observatory of the National Academy of Sciences of Ukraine (NASU) in Kiev, points out that it “might be hard for the [Russian] academy to receive the institutes since it is itself in the midst of reform”. However, herpetologist Oleg Kukushkin of the NASU’s Karadag Nature Reserve in southeast Crimea, claims that the NASU has been inefficient and bureaucratic, and says that research at the reserve would be best protected by moving to the Russian academy. “It would be the optimal and most logical decision,” he says.

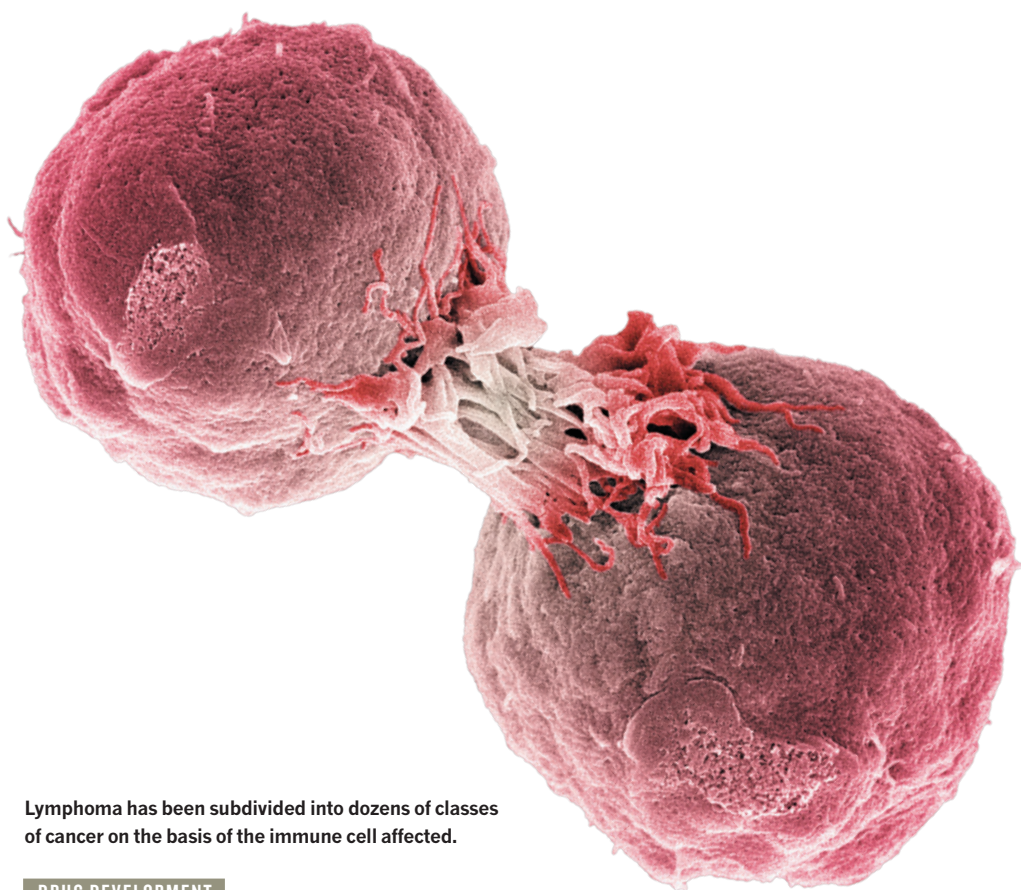
Elsewhere in Ukraine, the post-revolution euphoria and ambitions for rapid reforms of science governance have been dampened. And hopes are fading that the NASU, which comprises more than 500 members and runs about 170 research institutes and centres, would agree to modernizations that many say are crucial. This week, it postponed the planned elections of a new executive council until the autumn because of the political instability in Ukraine. NASU president Boris Paton, 95, who has been in office for 52 years, will now retain his role until next year.

Young researchers would like to see an end to this gerontocracy. Oleksandr Skorokhod, a PhD student at the Institute of Molecular Biology and Genetics in Kiev, is part of a newly organized network called the NASU Council of Young Scientists. Its members are lobbying effectively for systemic changes to improve the long-term health of Ukrainian science.

The network has also been included in the education ministry’s broad consultation process on the draft reforms. The reforms are being packaged into a new higher-education law, as well as into a revision to an existing science law. They include, for example, substituting the two Soviet-style higher degrees — candidate of science and doctor of science — with a PhD, which will be awarded by universities rather than the academy. The reforms would also impose Western-style standards for accrediting universities, and create an independent grants agency.

Shulga says that the reforms are being slowed down by the NASU because “most academy members are powerful and resistant to change”. Other scientists point out that resistance is encouraged by the generous lifelong stipend each member enjoys.

But Sergei Mosyakin, a corresponding member of the NASU and an ecologist at the M. G. Kholodny Institute of Botany in Kiev, says that the situation is not that simple. “The budget of the academy will be cut, so there will be no possibility of a real and productive large-scale reform. The general consensus is that the changes are inevitable, but they should be evolutionary.” ■



Lymphoma has been subdivided into dozens of classes of cancer on the basis of the immune cell affected.

DRUG DEVELOPMENT

Regulators adopt more orphan drugs

Agencies face rising applications for rare-disease therapies resulting from increasingly precise disease definitions.

BY SARA REARDON

Like a cancer cell, the definition of lymphoma has been dividing. Nearly 700,000 people in the United States have the disease, but over the past few decades, lymphoma has been sliced into several dozen subgroups on the basis of the type of immune cell that becomes cancerous. Drug companies have targeted these subtypes and have asked that they be counted as ‘orphan diseases’ — in the United States, this refers to diseases that have no cure and that affect fewer than 200,000 people, according to the definition given by the US Food and Drug Administration (FDA).

In 2013, the FDA granted orphan-drug designations to specific lymphoma treatments at least 21 times. And overall the programme — run by the Office of Orphan Products Development (OOPD) — designated

a record 260 drug applications as orphans last year, a 38% jump from 2012.

But as the trend towards precision medicine continues, some worry that the programme risks becoming a victim of its own prodigious success. The benefits given to developers through the scheme — such as market exclusivity — are helping to send the prices of orphan drugs sky high. And because drug applications for common diseases are stagnating, one of the FDA’s main revenue sources is under threat.

“Orphans are taking over the agency,” says Timothy Coté, a former OOPD director who now runs his own business, Coté Orphan Consulting, in Silver Spring, Maryland. In all, more than one-third of new drugs approved by the FDA last year were orphans.

The 1983 Orphan Drug Act was intended to give drug companies generous incentives to develop products that might not otherwise

STEVE GSCHMEISSNER/SPL

SOURCE: FDA/EMA

seem to be worthwhile investments. Companies typically apply for an orphan-drug designation early in the drug-development process. If an orphan drug is ultimately approved for the market, the FDA waives the US\$2.17-million 'user fee' that companies must pay to the FDA for new drugs. The companies also get tax credits for incurred clinical-trial costs, along with seven years of market exclusivity, during which time approvals for similar drugs are blocked.

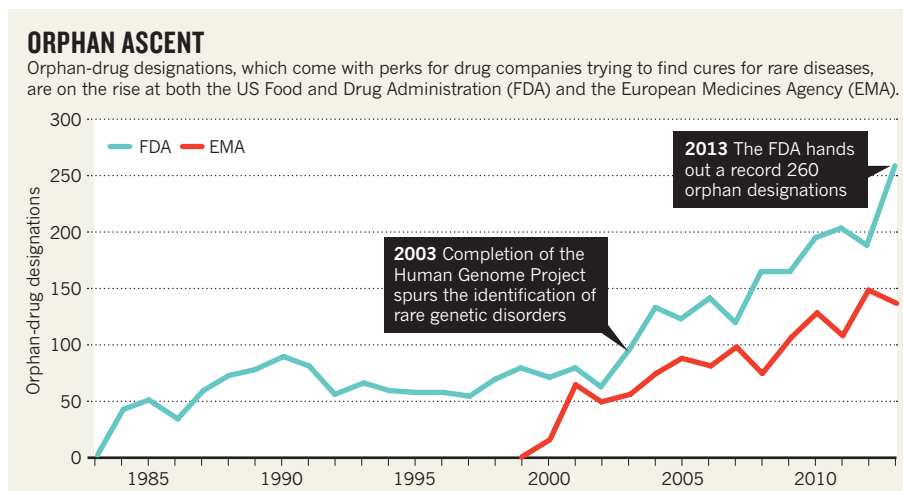
Europe has a similar scheme: in 1999, the European Medicines Agency created an orphan designation for diseases that affect fewer than 5 in 10,000 people in Europe. The programme, which among other perks grants 10 years of market exclusivity, has also seen a steady rise in orphan designations over the past few years (see 'Orphan ascent').

Companies are flocking to orphan drugs partly because of the difficulty in finding significantly better treatments for common diseases that already are well supplied with medicines. Dozens of treatments already exist for diabetes, for example, and many of them are quite effective. "The age of the blockbuster appears to be over," says Coté.

But drug developers are also finding that the market for some orphan drugs is bigger than one might think. New York City-based Pfizer, for example, is developing orphan treatments for sickle-cell disease. Although the disease is rare in the United States — where it affects fewer than 100,000 people — sickle-cell anaemia is common in Africa, and the company factors that into its calculations, says Alvin Shih, chief operating officer of Pfizer's rare-disease research unit in Cambridge, Massachusetts. He says that the company established the unit in 2010 after deciding that rare diseases would be a good investment.

Even when diseases are extremely rare, orphan drugs can still be profitable. In 2012, for instance, the FDA approved a drug called Kalydeco (ivacaftor; Vertex Pharmaceuticals, South Boston, Massachusetts), which targets one particular genetic mutation that causes cystic fibrosis, a disease that affects about 30,000 people in the United States. The drug works for only about 4% of patients with cystic fibrosis, and they pay as much as \$373,000 per year for it.

Some critics have expressed concerns that companies might artificially 'salami-slice' diseases. To protect against that, the FDA issued



a rule in June 2013 that clarified its position: applicants must provide scientifically plausible evidence for the uniqueness of a disease. But scientific advances provide ample opportunities to offer such justification. Geneticists have connected tiny mutations to rare syndromes; virologists have tracked the evolution of receptors on the surfaces of flu viruses; and biochemists have discovered the misfolded proteins that appear in diseases such as Alzheimer's.

In 1999, fewer than 1,000 diseases had a known mechanism, says Christopher Austin, who heads the National Center for Advancing Translational Sciences in Rockville, Maryland. Today, the mechanisms behind more than 5,000 diseases are known. "The common ones are becoming rare, the rare ones are becoming defined," Austin says.

The FDA's definition of an orphan drug reaches beyond medicines that treat very specific diseases. Drugs that effectively treat a small subgroup of patients with a common disorder — such as people for whom previous treatments for depression have failed — can also be designated as orphans, says Gayatri Rao, who heads the OOPD. Even an infection by bacteria with a certain type of antibiotic resistance could potentially get the designation.

These and different subtypes of a cancer are a far cry from the types of extremely rare genetic

disease that were the original motivation in establishing the orphan programmes — such as progeria, which causes premature ageing and affects fewer than 250 people in the world. "At the time, people didn't foresee oncology going in the direction it has," says Steven Bates, chief executive of the BioIndustry Association, an organization of UK pharmaceutical companies based in London.

Coté says that it is entirely conceivable that orphans might one day account for most of the FDA's drug approvals — and that could ultimately sap the agency of a major income stream. He notes that the \$2.17-million user fee that the FDA depends on to help fund its operations has already risen dramatically, up from \$573,500 just a decade ago, and he suspects that is partly because of the increasing number of orphan drugs that do not warrant any fee.

And the exorbitant cost of some orphan drugs is becoming a concern for insurers and national health-care systems in the European Union, says Bates. The UK National Institute for Health and Care Excellence, for instance, is considering ways to handle these expensive drugs, such as price ceilings or having companies justify their prices.

Eventually, says Rao, regulatory agencies might need to rethink their strategies. But for now, the FDA is glad that the orphan programme is so popular. "These are sort of long-term considerations, and we're not there yet," she says. "In the near future, the more orphan drugs we see, the better." ■

"Orphans are taking over the agency. The age of the blockbuster appears to be over."

MORE ONLINE

TOP NEWS



First synthetic eukaryotic chromosome created
go.nature.com/kgtlc

MORE NEWS

- 'Electronic skin' equipped with memory go.nature.com/4oxyes
- Mutations could explain low effectiveness of flu vaccine go.nature.com/slzvpr
- WHO doubles estimate of deaths from air pollution go.nature.com/kslybn

NATURE PODCAST

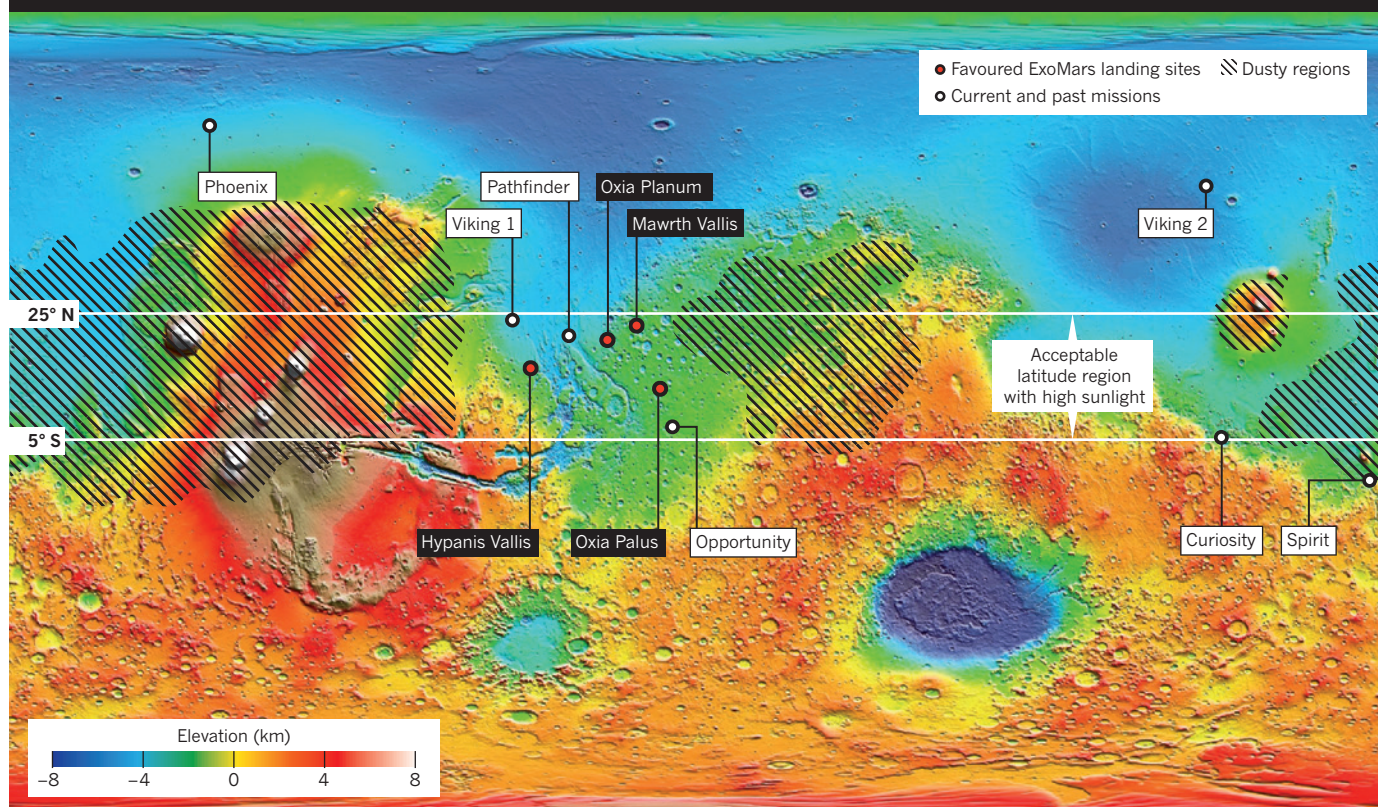


Gene expression in the prenatal brain; *Homo habilis*; and cancer immunotherapy
nature.com/nature/podcast

SOURCE: NASA/GODDARD SPACE FLIGHT CENTER

HOMING IN

European scientists have selected four favourable landing sites for the ExoMars rover. The sites must satisfy three engineering requirements: a low elevation, plentiful sunlight and little dust.



SPACE

ExoMars scientists narrow down landing sites

Facing engineering constraints, researchers propose four destinations for European rover.

BY ELIZABETH GIBNEY

Scientists have picked four potential landing sites for a European rover designed to search for life on Mars. The 300-kilogram ExoMars rover, part of a joint mission of the European Space Agency (ESA) and Russia's space agency, Roscosmos, is scheduled to land on the red planet in early 2019. Armed with a drill that can bore 2 metres into rock, the rover will search for preserved organic matter that, on the surface, might have been destroyed by harsh radiation.

Last week, 60 scientists met at ESA's European Space Astronomy Centre near Madrid to discuss eight potential sites. On 27 March, they cast an informal vote, and four sites emerged as favourites: Mawrth Vallis, Oxa Planum, Hypanis Vallis and Oxa Palus (see 'Homing in'). An expert working group tasked with

recommending a final landing site will now consider the proposals, before announcing a formal shortlist of three or four sites in June. Following detailed studies of the shortlisted sites, the panel will make a single recommendation to ESA and Roscosmos in late 2016.

Finding the right spot is crucial, says Matt Balme, a planetary scientist at the Open University in Milton Keynes, UK. "If we make the wrong choice, we could land somewhere unsafe — leading to total mission failure — or land somewhere where we have no chance of meeting the science objectives of the mission," he says.

Scientific and engineering constraints already rule out much of the planet. To have a chance of finding evidence of life, working-group scientists must pick a site that would have been watery more than 3.6 billion years ago, when water is thought to have been

more abundant. They also need a site where wet sediments were quickly buried and preserved in a stack of geological layers, says ESA's ExoMars project scientist Jorge Vago, who is based at the European Space Research and Technology Centre in Noordwijk, the Netherlands. The four leading sites fall into two categories: places where there is strong mineralogical evidence for wet sediments, and places where there is strong morphological evidence for river mechanisms that would have deposited fine sediments, which are best for trapping and preserving organic materials.

Mawrth Vallis and Oxa Planum, two vast plains that contain some of the planet's oldest rocks, fall into the first category. Minerals called phyllosilicates detected from orbit indicate that the rocks were once wet clays, and that they formed in a neutral pH environment that would have been favourable for life.

Meanwhile, images from orbit show that Oxia Palus once had a meandering river channel, and that Hypanis Vallis was once a delta. Both features would have buried organic matter quickly, leading to its preservation in sedimentary rock.

"Should we go where there is evidence of a wide-scale river environment, or go where the clays are?" asks John Bridges, a planetary scientist at the University of Leicester, UK, and a member of the site-selection working group. "There are good arguments both ways."

Jack Mustard, a geologist at Brown University in Providence, Rhode Island, and former chair of NASA's Mars advisory group, says that the best place to find signs of life will be somewhere with a long history of flowing water. "You want to be in a place where water would have been around long enough to have captured the spark of life, should it have ever started," he says. For this, he adds, Mawrth Vallis "stands out", because its chemistry suggests that the region held water for an extended time. The site narrowly missed out on being the destination for NASA's Curiosity rover, which landed in Gale Crater in 2012.

But before ExoMars can begin to test for life, it has to land safely. Responsibility for the landing system lies with Roscosmos, which signed up to the project last year after NASA dropped out in 2011 (see *Nature* <http://doi.org/bwd9hh>; 2011). Russia's expertise is built

"Should we go where there is evidence of a wide-scale river environment, or go where the clays are?"

mostly on its success with lunar landers in the 1960s and 70s. Vago says that the team will get to test its technology with a trial lander — the ExoMars Entry, Descent and Landing Demonstrator Module — in 2016, before the main mission.

The landing technology is already limiting the number of viable sites. ExoMars will use a heat shield and parachutes to pass through the thin atmosphere. It needs to travel through as much atmosphere as possible to slow it down sufficiently for a safe landing, so the landing zone must be low-lying. Furthermore, unlike the radioisotope-powered Curiosity, the rover will run solely on solar energy. That means

that only dust-free and sunny destinations, not too far from the equator, will do.

The landing window for ExoMars, given the uncertainties in its trajectory, is an ellipse 104 kilometres long and 19 kilometres wide — more than ten times the area of the window for Curiosity, which had the ability to land more precisely. This entire site must be mostly clear of large boulders and craters. Yet because the ExoMars rover will travel only a few kilometres in total, its path must contain scientifically interesting sites throughout. Over the next month, the working group will try to find the best landing ellipse for each site, says Vago.

NASA, for its part, will in May begin the process of picking a destination for its next Mars rover, scheduled for launch in 2020. Getting the data needed to evaluate sites for both ExoMars and the 2020 rover will be a race against time, says Mustard. NASA's Mars Reconnaissance Orbiter, which has done the lion's share of the high-resolution spectroscopy and imaging of the landing sites, is now nine years old, and starting to suffer from technical glitches. "The idea is to get the data before the pieces start falling off," says Mustard. ■

CLIMATE

El Niño tests forecasters

As hints emerge of a major weather event this year, poor data could thwart attempts to improve predictions.

BY JEFF TOLLEFSON

The first sign of a brewing El Niño weather pattern came in January, as trade winds that normally blow from the east reversed course near Papua New Guinea. Barrelling back across the tropical Pacific Ocean, they began to push warm water towards South America. Now climate scientists and forecasters are on high alert.

A major El Niño event — a periodic warming of waters in the eastern equatorial Pacific — could boost temperatures and scramble weather worldwide. The most recent major event, in 1997–98, was linked to thousands of deaths and tens of billions of dollars in damage from droughts, fires and floods across several continents. Yet more than 15 years later, forecasting the timing and intensity of El Niño remains tricky, with incremental improvements in climate models threatened by the partial collapse of an ocean-monitoring system that delivers the data to feed those models.

El Niño often emerges during the Northern Hemisphere summer and peaks around December; forecast models can do a reasonable

job of predicting its eventual strength by July, when the changes in ocean circulation that give rise to the weather pattern have become pronounced. But scientists are working feverishly to provide earlier forecasts, to allow governments more time to prepare for potentially devastating weather patterns.

In 1997, the emergence of a record-breaking El Niño caught scientists by surprise, despite hints in wind and sea surface temperature data (see 'Warming up'). The European Centre for Medium-Range Weather Forecasts in Reading, UK, was reluctant to issue early warnings because its forecast model was untested. And the model that had best predicted earlier El Niños — developed by scientists at Columbia University's Lamont–Doherty Earth Observatory in Palisades, New York — foresaw neutral conditions.

"We were in the early days in forecasting," says Michael McPhaden, an oceanographer at the US National Oceanic and Atmospheric Administration (NOAA) in Seattle, Washington. "Now we are much more systematic."

This year, NOAA issued its first forecast on 6 March, estimating a 50% chance that El Niño

will develop this summer. But that early projection, and others from weather agencies and research institutions around the world, comes with lots of uncertainty. Fickle tropical winds in spring can easily quash a brewing El Niño — or strengthen it.

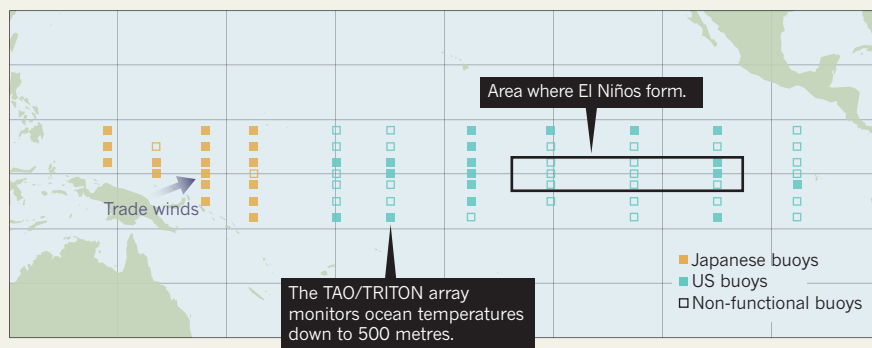
Researchers say that real progress in forecasting has come from systematically comparing the outputs of groups of models, with each simulation run under a range of possible climate conditions. "Combining these various predictions — doing some crowd-sourcing, if you will — tends to lead to more reliable predictions," says Gabriel Vecchi, a climate modeller at NOAA's Geophysical Fluid Dynamics Laboratory in Princeton, New Jersey. Averaging the results of several different forecasting models tends to cancel out flaws in an individual program, he says.

The sensitivity of the simulations is also increasing. Global climate models divide the planet into grids, like a computer screen divided into pixels, and represent climate variables such as temperature as averages in each cell. Modellers increase the resolution of their calculations by reducing the size of the cells.

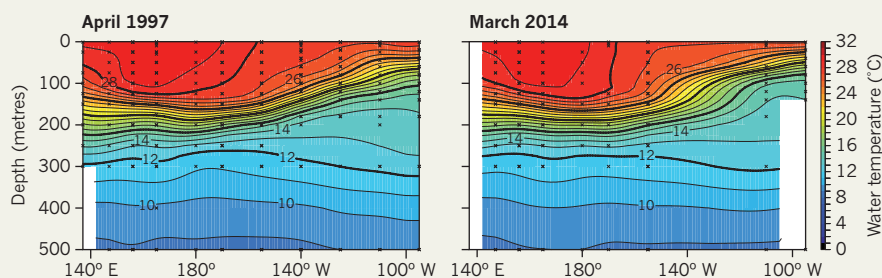
SOURCE: NOAA

WARMING UP

The US-Japanese TAO/TRITON array, a network of moored buoys in the Pacific Ocean, measures ocean temperature in the region where El Niño forms. US budget cuts have hampered buoy maintenance, causing the array to partially collapse. This could make tracking a brewing El Niño difficult.



Trade winds that normally blow towards the west briefly reversed course in January and again in February, pushing warm surface water towards South America, where El Niños take shape. Similar conditions in early 1997 helped to produce the most recent major El Niño event.



Vecchi's lab, for example, has shifted its model from a grid with cells 200 kilometres across to one with cells 50 kilometres across, thereby increasing the number of cells by a factor of 16. In theory, this allows for more-accurate representations of the microphysical processes and interactions that drive weather and ultimately climate. Forecasters also continuously fine-tune how they incorporate environmental-monitoring data and represent complex interactions between air and ocean circulation.

During the spring, when forecasting is most difficult, such improvements have given climate models based on physical processes a leg-up over less sophisticated statistical models, which compare the current weather-system trends to those of past years and essentially estimate the likelihood that history will repeat itself.

Almost all of this year's initial forecasts suggest that a moderate to severe El Niño or neutral conditions will emerge in coming months. None predicts El Niño's sister effect, La Niña, in which upwelling currents from the deep ocean bring cooler waters to the surface off the Pacific coast of South America.

In the coming weeks, scientists will watch to see if warm water continues to flow across the Pacific into the area off South America where El Niño forms. But in a potential blow

Fickle tropical winds can easily quash a brewing El Niño — or strengthen it.

to the ongoing effort to improve forecast accuracy, their ocean-temperature data will get progressively worse. A US-funded system of data-gathering buoys known as the Tropical Atmosphere Ocean (TAO) array has started to break down as a result of budget cuts that have hobbled its maintenance (see *Nature* <http://doi.org/q72>; 2014). NOAA has committed to restoring most of the system by the end of the year, but that aid will come months after crucial El Niño forecasts are issued. Scientists will be forced to supplement the buoy data with satellite observations of water temperature and sea level, which can serve as proxies for the depth of the wave of warm water.

The stakes are high. Since 1998, the eastern Pacific has been in a cold phase that is associated with La Niña-like conditions, but every 15–30 years, as part of a cycle known as the Pacific Decadal Oscillation, that trend flips. Kevin Trenberth, a climate scientist at the US National Center for Atmospheric Research in Boulder, Colorado, has theorized that a major El Niño could help to push the ocean back into a warm phase, which studies have linked to more frequent El Niños and more rapid global warming (see *Nature* **505**, 276–278; 2014).

But all of that depends on what happens as warm water washes across the Pacific in the next couple of months. “The system is primed,” says Trenberth. “Will it wimp out or really take off?” ■

Epigenomics starts to make its mark

Analysis of chemical patterns on DNA shows promise for explaining disease, but few results have yet been replicated.

BY EWEN CALLAWAY

Obesity may be written not only in the genes, but also on top of them.

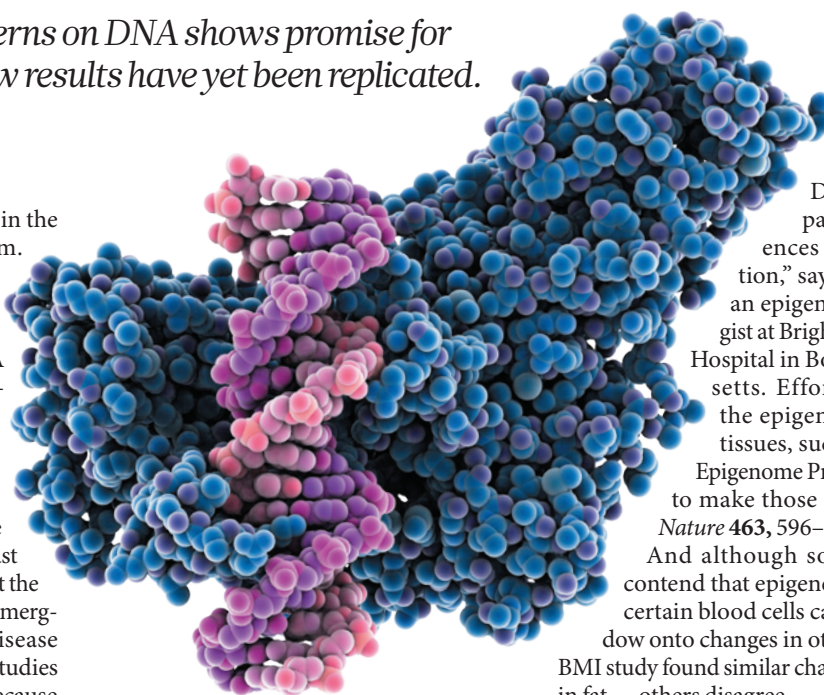
One of the largest studies so far to probe the human epigenome — the collective name for the patterns of chemical groups that adorn DNA sequences and influence their activity — has found some tags that are linked to differences in body mass index (BMI).

Chemical alterations in a gene thought to be involved in metabolism were identified in the blood and fat cells of more than 2,500 people¹, scientists reported last month. The work, led by Nilesch Samani at the University of Leicester, UK, is part of an emerging line of research that is probing disease through epigenome-wide association studies (EWAS). Those in the field hope that because many epigenetic changes are influenced by the environment, such studies will reveal mechanisms of disease that have remained elusive. However, many scientists remain sceptical.

EWAS take their inspiration from genome-wide association studies (GWAS), which compare genetic variations in ill and healthy people to identify variations associated with disease. EWAS do the same for the epigenome. They look at differences in the distribution of methyl groups at thousands of specific DNA nucleotides across the genome to identify arrangements that are common in a disease, or associated with variation in a trait.

Methylation — the addition of methyl groups — tends to suppress the activity of genes. It is important in development, when it helps to guide the differentiation of embryonic stem cells into specialized tissues by orchestrating the expression of genes. But it also occurs in response to environmental changes, and these gene modifications may be inherited. They may also contribute to conditions such as cancer and type 2 diabetes.

Because epigenetic alterations can occur over a person's lifetime, they could offer clues about the biology of diseases that genetics cannot. Many of the common genetic variants associated with disease that have been identified by GWAS have failed to explain fully why one person develops a condition and another does not. "Originally we were told that



A methyltransferase enzyme (blue) adds chemical groups to DNA (pink), altering gene expression.

common genetic variants would explain common diseases, and that's turned out largely not to be the case," says Andy Feinberg, director of the Center for Epigenetics at Johns Hopkins University in Baltimore, Maryland. "The environment is extremely important, and we pay almost no attention to it in biology."

But many scientists say that the field of EWAS, which is less than a decade old, has yet to show that it can yield real insight. "The problem with EWAS is that there's so much more that can confound an outcome compared with a GWAS," says John Gready, an epigeneticist at Albert Einstein College of Medicine in New York. For instance, epigenetic alterations vary drastically between cells of the same tissue, and it can be difficult to determine whether an alteration is a cause or consequence of a disease.

Accounting for differences between cells may be the field's biggest bugbear. Most EWAS have been conducted in blood cells, because blood is easy to collect. But it is also made up of many cell types, with varying epigenetic profiles. A study published in February², for instance, found that epigenetic signatures thought to result from ageing instead reflected the changing proportions of blood cell types with age.

"We have to distinguish how much of the differences between two individuals is due

to differences in DNA methylation pattern or to differences in cell composition," says Karin Michels, an epigenetic epidemiologist at Brigham and Women's Hospital in Boston, Massachusetts. Efforts to catalogue the epigenomes of myriad tissues, such as the Human Epigenome Project, should help to make those distinctions (see *Nature* **463**, 596–597; 2010).

And although some researchers contend that epigenetic alterations in certain blood cells can provide a window onto changes in other tissues — the BMI study found similar changes in blood and in fat — others disagree.

Stephan Beck, a medical genomicist at University College London, thinks that epigenomic epidemiology is at the same stage genomic epidemiology was at eight years ago, when most studies were small and rarely identified the same genetic variants for any one disease. That changed in 2007, when the Wellcome Trust Case Control Consortium, an effort by research groups to identify genetic variants linked to common diseases, set standards for GWAS that emphasized the importance of large numbers of patients and reproducibility.

Beck sees signs that EWAS are headed in the same direction. For example, in the past few years several studies have reproduced the finding that a gene called *AHRR* is epigenetically modified in the blood cells of adult smokers³ and their newborn children⁴. And some epigenome studies — including the BMI research¹ — now verify discoveries made in one group of patients in separate cohorts of the same study.

The pay-off for getting such studies right could be new therapies and insight into disease. "Epigenomics is a challenge," says Michels, "but I don't think we can afford to ignore it." ■

1. Dick, K. J. *et al.* *The Lancet* [http://dx.doi.org/10.1016/S0140-6736\(13\)62674-4](http://dx.doi.org/10.1016/S0140-6736(13)62674-4) (2014).
2. Jaffe, A. E. & Irizarry, R. A. *Genome Biol.* **15**, R31 (2014).
3. Philibert, R. A., Beach, S. R. H., Lei, M.-K. & Brody, G. H. *Clin. Epigenet.* **5**, 19 (2013).
4. Joubert, B. R. *et al.* *Environ. Health Perspect.* **120**, 1425–1431 (2012).

LAGUNA DESIGN/SPL

THE KILLER WITHIN

The immune system can be a powerful weapon against cancer — but researchers are still grappling with how to control it.

The first tumour was a small melanoma on the left side of attorney Mark Gorman's neck. Doctors removed it, and assured him that the cancer was gone.

But eight years later, in 1998, a physician felt Gorman's abdomen during a routine physical examination, arched an eyebrow, and asked if he had become a heavy drinker. The melanoma had spread to Gorman's liver, seeding an inoperable beast of a tumour that wrapped around the inferior vena cava carrying blood to his heart.

People with advanced melanoma typically live for just six to ten months after diagnosis. But Gorman, then 49, had little patience for the doctors advising him to get his affairs in order. When his sister told him about a drug called interleukin-2 (IL-2) that was being used together with chemotherapy against melanoma at a hospital in Colorado, he travelled from his home in Silver Spring, Maryland, to give it a try.

IL-2 is a protein produced by white blood cells called T cells during an immune response. Taking high doses of it sends T cells into overdrive, making them more likely to recognize and attack cancer cells. Gorman was treated, and remains cancer-free 15 years later. "Some doctors say my immune system is really smart," he says. "I just know I'm lucky."

The drug that saved Gorman's life was the first treatment approved by the US Food and Drug Administration (FDA) to fire up the immune system's response to cancer — a technique known as immunotherapy. After that 1992 approval, researchers and pharmaceutical companies spent years trying to develop new immunotherapies that could produce success stories like Gorman's. But those attempts failed to live up to their promise in the clinic, leading to decades of frustration.

Now the tide seems to be turning. Clinical-trial successes in the past five years suggest that a new generation of approaches has potential against several forms of cancer that resist conventional treatments. Some analysts predict that in the next ten years, immunotherapies will be used for 60% of people with advanced cancer, and will comprise a US\$35-billion market. "It is kind of crazy," says Cary Pfeffer, chief executive of Jounce Therapeutics, a company specializing in cancer immunotherapy in Cambridge, Massachusetts. "This field has become so crowded. It's frenzied."

But the sobering experience with earlier drugs has made many researchers and clinicians cautious. Despite its potential for miracles,

BY HEIDI LEDFORD

IL-2 produces complete remission in only around 6% of people with melanoma. The treatment kills as many as 2% of recipients. Researchers are now racing to find ways

to boost the number of patients who respond to immunotherapy and to reduce the dangerous side effects. "The good news — and the bad news — is that the immune system is incredibly powerful," says Robert Tepper, chief medical officer at Jounce.

CHECKMATE

Cancer immunotherapy was born in 1891, when a New York surgeon named William Coley began injecting bacteria into patients' tumours in the hope of triggering an immune response to the infection that would also attack the tumour. Physicians before him had noted mysterious and rare cancer remissions following infections, and Coley was eager to harness that therapeutic power.

It would not be so simple. Tumours wield many defences against the immune system's most powerful cancer-fighting weapon: T cells that hunt out and eliminate problem cells. Cancer cells disguise themselves and make it difficult for T cells to find them. Tumours also fend off immune attack by expressing proteins that suppress T cells in the surrounding environment.

For decades, researchers chased the possibility of a vaccine that would alert the immune system to cancer cells. But those efforts have largely failed: the only FDA-approved therapeutic cancer vaccine is a complicated and costly therapy for prostate cancer. Whether it provides a significant benefit to patients is a matter of debate.

The field turned a corner in 2011, when the FDA approved a new kind of immunotherapeutic drug. Yervoy (ipilimumab) binds to and blocks a 'checkpoint' protein called CTLA-4, which normally acts as a brake on the immune system by preventing T-cell activation. Checkpoint proteins keep the cells in check so that they do not attack normal tissue. When Yervoy releases the brake, T cells are free to destroy tumours.

Like IL-2, Yervoy can bring long-lasting responses. Some participants in the original trials have been in remission for 13 years, says James Allison, a cancer immunologist at the University of Texas MD Anderson Cancer Center in Houston. But those clinical cures occur in just a small fraction — about 8% — of patients. And although Yervoy can rouse T cells to battle against

➔ NATURE.COM

To hear more about immunotherapy, see: go.nature.com/nqaxb5

cancer, sometimes the cells attack healthy tissue, too. Of the 540 people who took Yervoy in the largest trial, up to 15% experienced serious side effects and seven died of immune-related events. Some oncologists prefer to steer clear of the drug, says Suzanne Topalian, a melanoma researcher at Johns Hopkins University School of Medicine in Baltimore, Maryland.

Still, the promising aspects of Yervoy established the potential of checkpoint inhibitors — drugs that block checkpoint proteins — and that has prompted researchers to look at other potential target proteins. By the time Yervoy was approved, some investigators had begun to focus on PD-1, a checkpoint protein that some cancers use to deactivate the phalanx of T cells that surrounds the tumour.

Because PD-1 interacts directly with cancer cells, unlike CTLA-4, its inhibitors have the potential to be more potent and less toxic. Early clinical trials suggest that this is the case. A leading PD-1 inhibitor — nivolumab, made by New York's Bristol-Myers Squibb — shrinks tumours in 28% of people with advanced melanoma. The FDA is expected to issue a decision on whether to approve it by early 2015, if not sooner.

Hopes are high that, although there are some side effects, the new drugs will be less toxic than Yervoy. Some people notice no problems at all. “Many patients say, ‘Doc, are you even giving me anything?’” says Antoni Ribas, a melanoma specialist at the University of California, Los Angeles, who has participated in trials of PD-1 inhibitors. “Then the tumours start disappearing, and they know.”

Researchers want to push immunotherapies even further. “We wish response rates were higher than what we currently have,” says Michael Postow, an oncologist and cancer researcher at Memorial Sloan Kettering Cancer Center in New York. Inhibitors of other checkpoint proteins are trickling into clinical testing

and clinicians may one day match patients with the inhibitors most likely to act on the proteins expressed by their own cancer cells.

For other patients, the challenge may be in attracting T cells to the tumour in the first place. PD-1 inhibitors do not accomplish this — they simply remove the shackles from T cells already amassed at the tumour's edge, says Daniel Chen, head of immunotherapy development at Genentech in South San Francisco, California, a subsidiary of the Swiss pharmaceutical giant Roche. “Some patients just seem to have no existing immune response to start with,” he adds. “So then we need to add something that will generate that response.”

BETTER TOGETHER

The key to attracting T cells is to create an ‘inflamed’ tumour using combinations of therapies, says Postow. Yervoy and PD-1 inhibitors are already in clinical trials with each other and a range of other treatments intended to alert T cells to the cancer. Radiation, for example, breaks open cancer cells and releases antigens, molecules that can trigger immune responses. In another approach, researchers alert a patient's immune system with experimental cancer vaccines containing proteins that are overexpressed by tumour cells. “The future is clearly combination therapy,” says Anthony Marucci, chief executive of Celldex Therapeutics in Hampton, New Jersey.

Eventually, checkpoint inhibitors could also be combined with a form of immunotherapy called adoptive T-cell transfer. This is a personalized treatment in which physicians isolate T cells from patients and select those that react to cancer. They then multiply the T cells and stimulate them with molecules such as IL-2 before injecting them back into the bloodstream. Trials of this method led by tumour immunologist Steven Rosenberg at the National Cancer Institute in Bethesda, Maryland, have shrunk

**“THE GOOD NEWS —
AND THE BAD NEWS —
IS THAT THE IMMUNE
SYSTEM IS INCREDIBLY
POWERFUL.”**

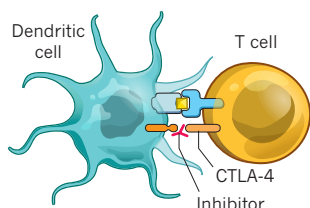


IMMUNE BOOST

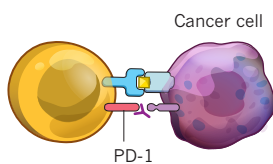
Several methods are showing promise in helping immune sentinels called T cells to attack cancer.

CHECKPOINT INHIBITOR DRUGS

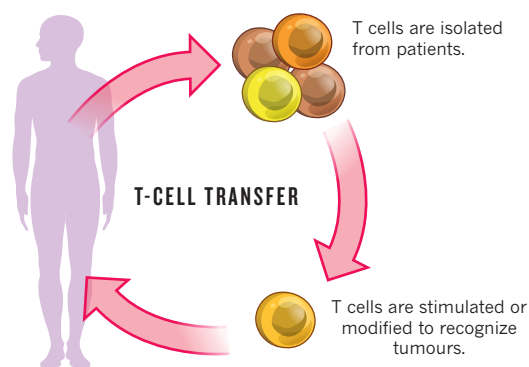
'Checkpoint' proteins block T-cell activity. Inhibitor drugs can release the brakes on T cells at different stages.



The CTLA-4 checkpoint protein prevents dendritic cells from priming T cells to recognize tumours. Inhibitor drugs block the checkpoint.



The PD-1 checkpoint protein prevents T cells from attacking cancer cells. The inhibitor drug allows T cells to act.



tumours in more than half of people with advanced melanoma receiving the treatment, with 20% experiencing complete remission.

A newer form of T-cell transfer promises to broaden its reach to other cancers, by engineering extracted T cells to express an artificial tumour-targeting receptor called a chimaeric antigen receptor (see 'Immune boost'). A trial using T cells engineered to target B cells wiped out cancer in 14 of 16 people with acute leukaemia (M. L. Davila *et al. Sci. Transl. Med.* **6**, 224ra25; 2014).

But technical challenges have limited the spread of T-cell transfer therapies. Only a handful of academic medical centres have performed the procedure so far. "After our initial results, we were besieged with melanoma patients," says Rosenberg. "We couldn't possibly treat all the patients sent to us."

Since those early days, researchers have simplified and standardized protocols. That, plus the remarkable results in leukaemia, has lured industry investors. Novartis, based in Basel, Switzerland, has bought a facility in New Jersey to process T cells extracted from patients around the United States. The facility will be key to the company's plans to expand its clinical trials to more sites this year. Smaller firms are following suit. In early 2015, Kite Pharmaceuticals in Santa Monica, California, hopes to launch a multicentre trial of adoptive T-cell transfer in a form of lymphoma that kills around 37% of patients within five years of diagnosis.

THE TRUE TARGET

Another big challenge for adoptive T-cell transfer is to broaden its reach by finding new molecular targets that will guide T cells to specific tumour types while sparing healthy cells. The approach works well in leukaemia and other cancers that affect B cells, another class of white blood cell, because researchers can engineer T cells to target a protein called CD19, which is found only on B cells. Although the treatment wipes out healthy B cells in addition to the cancerous ones, patients can tolerate that side effect relatively easily. But finding a similar target for

solid tumours, which are less uniform than liquid tumours, has been difficult. "It's a major limiting step," says Ribas. "We're all excited about CD19, but it's not clear what the next target will be."

Researchers are mining growing databases of gene expression to find the best candidates. But firing up immune responses to specific proteins can be dangerous: a few years ago, four patients died in trials of T cells engineered to attack cells expressing a protein called MAGE-A3. This protein is expressed only in embryos and in some cancer cells in adults, so it seemed an ideal target. But researchers later learned that the T cells attacked similar proteins present in the heart and brain. "These T cells are professional killers," says Arie Beldegrun, chief executive at Kite. "If their target is expressed even in minute quantities on normal cells, these super killers are going to find those cells and destroy them."

In response to the deaths, ImmunoCore, an immunotherapy company based in Abingdon, UK, developed new bioinformatic methods to search for signs that any possible T-cell target could be expressed in normal tissue. The company also began to do its initial safety testing in three-dimensional cell cultures that better reflected the cells' natural environment. This approach has led to a collection of more than 20 potential targets for various cancers. Michel Sadelain, a cancer geneticist at Memorial Sloan Kettering, hopes to engineer T cells that target two proteins, both of which would have to be expressed on a cell for the T cells to destroy it. The idea, he says, is that the chance that a healthy cell will have both targets on its surface will be slim.

Finding more targets could help immunotherapy to reach more types of cancer. So far, researchers have focused on melanoma and kidney cancer because they responded best to immunotherapies in early trials, and are thought to be particularly visible to the immune system.

Rosenberg says he is working on 11 clinical trials testing adoptive T-cell therapies against a variety of cancers, including a particularly lethal and rare form called mesothelioma. The door to much wider applications for cancer immunotherapies opened in 2012, when results showed that the checkpoint inhibitor nivolumab shrank tumours in 18% of people with certain types of advanced lung cancer (S. L. Topalian *et al. N. Engl. J. Med.* **366**, 2443–2454; 2012). Because lung cancer is one of the world's most prevalent forms of cancer, the results raised hopes that immunotherapy could make a sizeable dent in cancer deaths. "This was a cancer that we thought was not immunogenic," says Ribas, who notes that both Yervoy and IL-2 failed to shrink lung-cancer tumours. "We thought immunotherapy wouldn't have a chance."

Some cancers, including liver cancer, may still pose a challenge to immunotherapy approaches, says Lisa Butterfield, a cancer researcher at the University of Pittsburgh in Pennsylvania. The liver processes pathogens and antigens in the blood, and the immune system is carefully controlled there to avoid prompting reactions that would target an individual's normal cells. Breast, colorectal, pancreatic and ovarian cancers are also particularly adept at suppressing immune cells. Combination therapies may provide a way around these limitations, she says.

Combination therapies may also be the salvation of the cancer-vaccine concept. Although the vaccines tested thus far have fared poorly, they may work synergistically with other immunotherapies, says Willem Overwijk, a cancer researcher at MD Anderson.

After so many years of disappointing results, the growing excitement over immunotherapy has surprised many cancer researchers and families touched by the disease. Since his own remarkable recovery, Gorman has mourned again and again as friends he made at melanoma support groups succumbed. Then, a few years ago, he had a new experience: a close friend was given Yervoy, and went into full remission.

As for his own melanoma, Gorman goes for scans to look for new tumours every two years. In February, he noted that it might be time to schedule his next set of scans. But he wasn't sure — he had stopped fearing his cancer's return years ago. "I'm a cool cucumber now," he says. "My immune system has it under control." ■

Heidi Ledford reports for Nature from Boston, Massachusetts.

POLAR STAR

After years of work in the Antarctic, John Kovac and his team have captured strong evidence for a long-held theory about the Universe's birth.

BY RON COWEN

The taxi ride across Cambridge normally takes just 15 minutes. But on 10 March, astronomer John Kovac had a momentous secret to share and wanted to avoid being spotted by reporters or scientific rivals. So he left his office at the Harvard-Smithsonian Center for Astrophysics a few minutes early, and directed his driver to drop him outside the Center for Theoretical Physics at the Massachusetts Institute of Technology (MIT). That left him time to walk around the back of the building and climb a little-used staircase that led straight to the third-floor office of cosmologist Alan Guth.

Back in 1980, Guth had proposed an idea that was both startling and appealing. During the first tiny fraction of a second after the Big Bang, he had theorized, the Universe underwent 'inflation': a process of ultra-rapid expansion that took it from subatomic size to a scale so vast that no one will ever see it all. Because the inflation hypothesis posited that far-flung regions of the Universe had started off close together, it solved several enduring cosmological puzzles, including why distant reaches of the Universe look almost identical. Indeed, most cosmologists believed that inflation, or something very much like it, must have happened. Yet for more than three decades, the theory had lacked definitive proof.

Now, Kovac told Guth, proof seemed to be in hand. Kovac was principal investigator of a team that had spent 2010 to 2012 monitoring the skies over the South Pole with an ultra-sensitive microwave receiver known as BICEP2. By outfitting BICEP2 with the microwave equivalent of polarizing sunglasses, they had been able to detect subtle patterns in the microwave afterglow of the Big Bang. These patterns were the faint imprint of gravitational

waves — tremors in the fabric of space-time generated during inflation. And now, after painstakingly checking and rechecking those measurements, the team was set to make the results public in exactly one week's time.

The distribution and magnitude of the gravitational waves, Kovac told Guth, were just as predicted by the theory of inflation.

Guth grilled Kovac for an hour and a half, going over the team's draft paper line by line to verify the results. At the end of the meeting, Guth was convinced. "This is a wonderful result," he said later — an "incredibly strong piece of evidence for inflation."

The news was announced to the world in a headline-making press briefing at the Center for Astrophysics on 17 March. Everyone involved was well aware that the finding came from just one group, and still needed independent confirmation. Nonetheless, the thoughts of many in the field were voiced by Marc Kamionkowski, a cosmologist at Johns Hopkins University in Baltimore, Maryland, who was one of the first to predict the gravitational-wave imprint detected by BICEP2. After the briefing, he told reporters: "To me, this is as Nobel-prize-worthy as it gets."

But for the 43-year-old Kovac, one of the most gratifying aspects of the day was that two of the founders of inflationary cosmology — the 67-year-old Guth and 66-year-old Andrei Linde of Stanford University in California — were watching from the audience. "It's a rare thing in science that the originators of the theory are around when testable consequences are actually searched for and ultimately found," Kovac says. That is why he had taken time away from the team's feverish pre-announcement

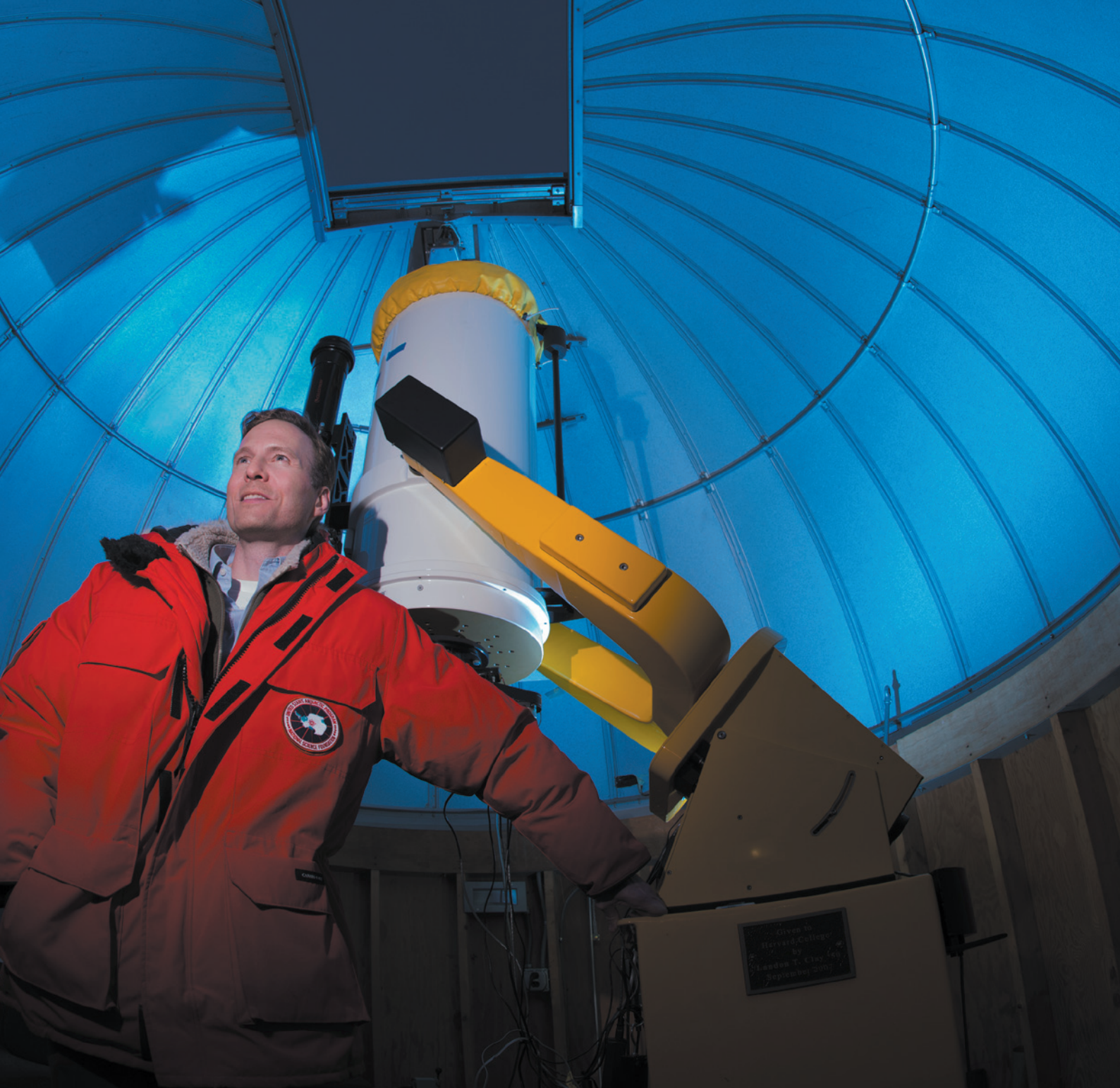
preparations to make the clandestine trip to see Guth at MIT. "We both realized it would be an important moment," he says.

BACK TO THE BEGINNING

Mention the prospect of a Nobel prize to Kovac, however, and he politely but firmly changes the subject. "I just think about the work," he says. It is a subject he talks about earnestly and methodically, while fixing his listener with a piercing blue-eyed stare.

It took that kind of intensity to detect the imprint of gravitational waves in the Big Bang's afterglow — known more formally as

RICK FRIEDMAN



the cosmic microwave background (CMB). The BICEP2 researchers had to measure temperature variations in the CMB as small as one ten-millionth of a kelvin. They had to detect the ever-so-slight polarization in the radiation, and how that varied with position. They had to isolate 'B modes': swirling patterns in the polarization that can be caused by gravitational waves. And they had to be extremely careful to make sure that the B modes they saw were actually in the CMB, and not caused by other things such as interstellar dust.

"It's been an emotional roller coaster," says Kovac, thinking back over his team's efforts

to determine that their signal was real. They had to be especially rigorous when they realized that the strength of their B-mode signal was about twice that extrapolated from non-polarization results reported last year from the European Space Agency's Planck spacecraft. The BICEP2 team eventually decided to go with its own data — but the discrepancy is still not fully understood.

Adding to the pressure was the threat of being scooped by several other teams that were racing to find the polarization signal. Kovac and his collaborators had to keep their results secret even from their close companions

working on the South Pole Telescope, which stands just metres from BICEP2. "We eat meals with them all the time," says Kovac. "We're friends. We party together." In fact, the team working on that telescope is led by Kovac's former thesis adviser, John Carlstrom of the University of Chicago in Illinois. "I've been dying to talk with John about this," says Kovac. "But professionally, we all know how these things work."

Despite the urgency, Kovac was a perfectionist about the analysis. But finally, he called a team meeting at the South Pole in early December 2013, laying out all the tests the data had passed and the milestones still to be

achieved. If the data held up, he told his group, the team was ready to publish. It was an intense meeting because so much was at stake, recalls Kovac. But, he says, “my role in this process has been to remain calm at all times.”

Kovac learned this cool, systematic approach to problems from his father Michael, a former dean of engineering at the University of South Florida in Tampa, who died two years ago. “My dad was an amazing guy and full of wisdom on how to lead teams, how to organize efforts in science,” says Kovac; he still has trouble talking about his father without choking up.

It was Michael Kovac who guided John into science. From the time he could talk, Kovac says, he was always asking questions. “My dad was able to feed that by answering every question that I asked him about how the world works in a way that explained to me what I wanted to know and led me to another question.”

When Kovac was nine, he became fascinated by the integrated circuits his father had been studying. “John never met a machine he didn’t want to take apart and find out how it worked,” recalls his mother, Midge. He built a simple calculator from parts his father had brought home from work. Then at the age of ten, he graduated to building a half-a-million-volt van de Graaff generator. “It was fun to zap all my friends at birthday parties,” he says.

Another strong influence was Lottie Peterson, a science teacher at his elementary school. Peterson had studied at the University of Chicago during the tenure of Enrico Fermi, the Nobel-prizewinning particle physicist. “She was able to tell me stories that captivated me about what it really meant to do physics at those high levels,” says Kovac. Peterson also gave Kovac’s family a telescope that he set up behind his house.

By the time Kovac entered secondary school, he was determined to be a scientist. After reading every cosmology book in the school library within the first year, he asked for more. Among them was Steven Weinberg’s 1977 account of the moments that followed the Big Bang, *The First Three Minutes: A Modern View of the Origin of the Universe* (Basic Books). Here, Kovac encountered Weinberg’s description of the CMB: “a diffuse background of radio static left over from near the beginning of the universe.”

Kovac was hooked on exploring the CMB and the clues it held. “As a kid, it seemed clear to me that this was the coolest thing in all of science — there are no bigger questions.”

Kovac chose to go to Princeton University in New Jersey in part because some of the major players in the field of CMB astronomy were there. By a stroke of luck, he says, he was assigned a work-study job with one of those researchers, astronomer David Wilkinson, in a group that was planning to build a telescope at the South Pole to search the apparently uniform CMB for regions that were ever so slightly hotter or colder. These temperature variations



John Kovac and Sarah Kernasovskiy work on part of the Keck Array at the South Pole.

would signal the existence of fluctuations in the density of the rapidly cooling masses of hydrogen and helium that came out of the Big Bang. Measuring them was tantamount to seeing the ‘seeds’ that would eventually contract — as a result of gravity — to form the galaxies and clusters of galaxies seen today.

This prospect so captivated Kovac that he took a year out from university to join the team in Antarctica for the austral summer of 1990–91. Wilkinson’s group was soon beaten to the discovery of the first temperature fluctuations in the CMB by the team analysing data from NASA’s Cosmic Background Explorer

“AS A KID, IT SEEMED CLEAR TO ME THAT THIS WAS THE COOLEST THING IN ALL OF SCIENCE.”

satellite. But within a year, the Princeton group’s South Pole instrument detected the fluctuations, too, and Kovac was hooked once again. In the years since, he has made a further 22 visits to the South Pole, often stopping over in New Zealand en route to indulge his hobby of mountain-climbing in the nation’s Southern Alps. On one occasion, he stayed at the South Pole for the entire southern winter, a nine-month interval when planes are typically not permitted to fly to the Antarctic because of the dangers posed by the extreme cold.

Kovac is the only principal investigator in the field of CMB astronomy who has ‘wintered over’, says BICEP2 telescope engineer Steffen

Richter. And because of that experience, Richter adds, “he knows the telescope down to the last screw; whatever the problem is, if you get his attention and he starts focusing on it, you can solve it with him in very little time.”

During his years as a graduate student under Carlstrom, Kovac’s Antarctic trips included work on the Degree Angular Scale Interferometer: an array of radio telescopes that the group used to make the first detection of CMB polarization in 2002. “I think the world of John,” says Carlstrom. “He never really needed advising.”

Later, as a postdoc and then a senior fellow at the California Institute of Technology in Pasadena, Kovac worked in Andrew Lange’s laboratory on highly sensitive polarization detectors for the QUAD and BICEP1 radio telescopes, which were also based at the South Pole. “Andrew was an inspiration and a close friend,” says Kovac. “He entrusted me with a huge amount of responsibility, encouraging me to take charge of the deployment and operation of the BICEP1 telescope and then to step into the role of leader of BICEP2.”

On his bookshelf, Kovac keeps a picture of the late Caltech astrophysicist, who in 2010 lost his battle with depression and committed suicide. After Lange’s memorial service, Kovac had a meeting with the three other key researchers on the BICEP2 project — Clement Pryke of the University of Minnesota in Minneapolis, Jamie Bock of NASA’s Jet Propulsion Laboratory in Pasadena and Chao-Lin Kuo of Stanford. The four agreed to take an equal share in running the team’s South Pole programme, which has now upgraded BICEP2 into a five times more sensitive detector known as the Keck Array, and which will next year add an equally sensitive telescope called BICEP3 that will measure the CMB polarization at a different wavelength. Bock believes that such shared leadership is unique among CMB projects. “I feel our decisions are always better than what one single person would initially propose,” he says, adding that the unusual arrangement “works for us because we respect and trust each other.” Indeed, at the 17 March press briefing, all four scientists took turns presenting the BICEP2 findings.

The next day, at MIT, Linde addressed a packed lecture hall, giving the first talk on the theoretical implications of the BICEP2 results. Afterwards, Guth reminded the audience that just as the theory of inflation rests on the shoulders of others such as Newton and Einstein, the experimental techniques used by the BICEP2 team members depended on great developments in technology made by those who came before them.

Then, as cups of bubbly cider were poured, Guth proposed a toast: “To the power of scientific reasoning!” Kovac and the rest of the audience cheered. ■

Ron Cowen is a freelance writer in Silver Spring, Maryland.

COMMENT

SUSTAINABILITY Citizens and businesses must track governments' progress **p.33**

POLICY End the bureaucracy that is holding back science in India **p.36**

THEATRE Shakespeare's world was steeped in practical discovery **p.39**

ENVIRONMENT James Lovelock on surprisingly optimistic form **p.41**



Fifty years after *Homo habilis*

Bernard Wood explains why the announcement of 'handy man' in April 1964 threw the field of hominin evolution into a turmoil that continues to this day.

Half a century ago, the British–Kenyan palaeoanthropologist Louis Leakey and his colleagues made a controversial proposal: a collection of fossils from the Great Rift Valley in Tanzania belonged to a new species within our own genus¹. The announcement of *Homo habilis* was a turning point in palaeoanthropology. It shifted the search for the first humans from Asia to Africa and began a controversy that endures

to this day. Even with all the fossil evidence and analytical techniques from the past 50 years, a convincing hypothesis for the origin of *Homo* remains elusive.

In 1960, the twig of the tree of life that contains hominins — modern humans, their ancestors, and other forms more closely related to humans than to chimpanzees and bonobos — looked remarkably straightforward. At its base was *Australopithecus*,

The foot of 'handy man', *Homo habilis*.

the apeman that palaeoanthropologists had been recovering in southern Africa since the 1920s. This, the thinking went, was replaced by the taller, larger-brained *Homo erectus* from Asia, which spread to Europe and evolved into Neanderthals, which evolved into *Homo sapiens*. But what lay between the australopiths and *H. erectus*, the first known human?

BETTING ON AFRICA

Until the 1960s, *H. erectus* had been found only in Asia. But when primitive stone-chopping tools were uncovered at Olduvai Gorge in Tanzania, Leakey became convinced that this is where he would find the earliest stone-tool makers, who he assumed would belong to our genus. Maybe, like the australopiths, our human ancestors also originated in Africa.

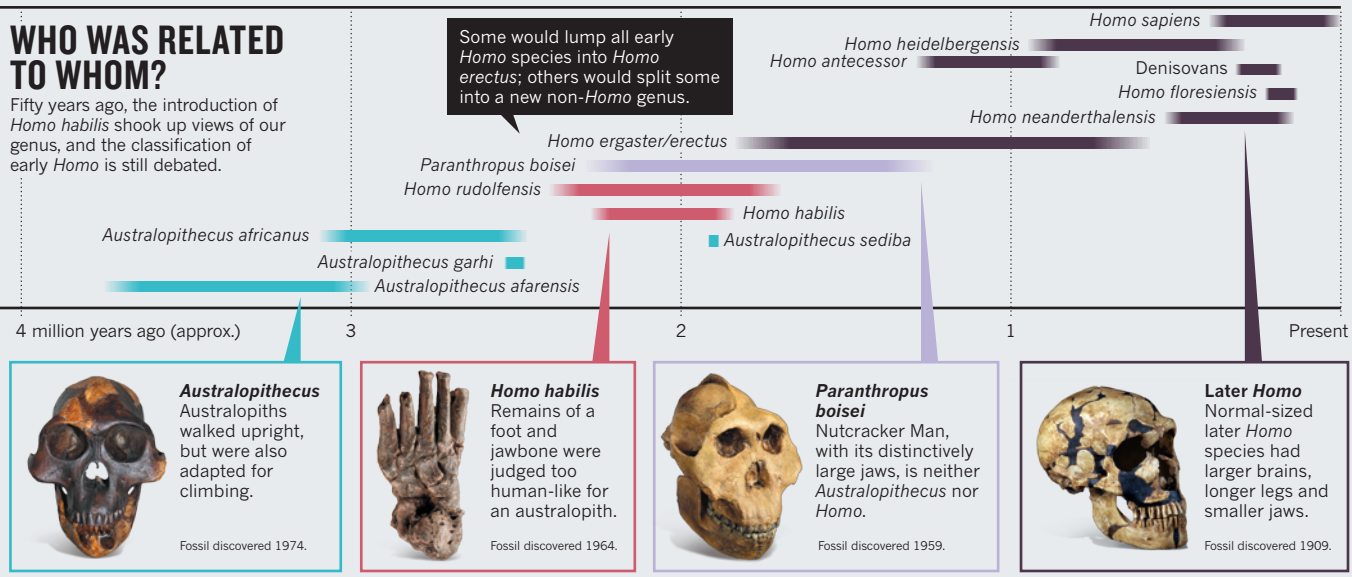
In 1931, Leakey began intensive prospecting and excavation at Olduvai Gorge, 33 years before he announced the new human species. Now tourists travel to Olduvai on paved roads in air-conditioned buses; in the 1930s in the rainy season, the journey from Nairobi could take weeks. The ravines at Olduvai offered unparalleled access to ancient strata, but fieldwork was no picnic in the park. Water was often scarce. Leakey and his team had to learn to share Olduvai with all of the wild animals that lived there, lions included.

They found the first trace of the potential toolmaker, two hominin teeth, in 1955. But these were milk teeth, which are not as easy to link to taxa as permanent teeth. The team's persistence was rewarded in 1959, when archaeologist Mary Leakey, Louis's wife, recovered the cranium of a young adult. The specimen still boggles the mind because it is so strange: its small brain, large face, tiny canines and massive, thumbnail-sized chewing teeth were not at all like those of *H. erectus*. Its big molars earned it the nickname 'Nutcracker Man'.

Because Nutcracker Man was found in the same layers as the stone tools, the Leakeys assumed that it was the toolmaker, despite its odd appearance. But when Louis announced the discovery, he was not tempted to expand the definition of *Homo*. That would have eliminated any meaningful distinction between humans and australopiths. Instead he erected a new genus and species, *Zinjanthropus boisei* (now called *Paranthropus* ▶

WHO WAS RELATED TO WHOM?

Fifty years ago, the introduction of *Homo habilis* shook up views of our genus, and the classification of early *Homo* is still debated.



► *boisei*), to accommodate it (see 'Who was related to whom?').

In 1960, Jonathan Leakey, Louis and Mary's eldest son, found the lower jaw and the top of the head of a juvenile hominin. Dubbed Johnny's Child, it very definitely did not belong to the same species as 'Zinj', and the Leakeys began to suspect that it was the real toolmaker.

Phillip Tobias, a palaeoanthropologist known for his work in South Africa, had already been recruited to analyse Zinj, so the Leakeys turned to him to analyse the juvenile cranium. John Napier, a specialist in hand anatomy (as well as sleight-of-hand magic tricks) was recruited to examine wrist and hand bones found with the skull.

An adult foot was excavated along with Johnny's Child, and three years later, a cranium with both the upper and lower jaw was uncovered, as was a very fragmented cranium with well-preserved teeth. Napier had already convinced himself that the juvenile hand bones were like those of modern humans. My PhD supervisor, Michael Day at the University of London had come to the same conclusion about the foot. And Tobias was certain that neither the long crowns of the chewing teeth in the lower jaw nor the large brain case could belong to the australopithecids known from southern Africa.

HANDY HYPOTHESES

Thus, in a paper published in *Nature* in April 1964 (ref. 1), Louis, Tobias and Napier made the case for adding the 'handy man' to the genus *Homo* as *H. habilis*. They argued that the Olduvai fossils met three key criteria set out in an influential 1955 definition of *Homo*²: an upright posture, a bipedal gait and the dexterity to fashion primitive stone tools. The team had to relax a brain-size criterion to accommodate the smaller brain

(around 600 cubic centimetres) of *H. habilis*.

The proposal was met with considerable scepticism. Some thought that the fossils were too similar to *Australopithecus africanus* to justify a new species. John Robinson, a leading authority on australopithecids, suggested that *H. habilis* was a mix of earlier *A. africanus* and later *H. erectus* bones. Other researchers agreed that the species was new. Very few accepted that it was the earliest human.

Subsequent finds shaped the debate. A crushed cranium (dubbed 'Twiggy') from the lowest strata at Olduvai nixed Robinson's argument that *H. habilis* was a mix of an australopithec and *H. erectus*. Another skeleton indicated that *H. habilis* had a stronger and relatively longer (or more ape-like) upper limb than did *H. erectus* and its ilk.

A handful of additional specimens from Ethiopia to South Africa have since been added to *H. habilis*; the biggest contribution to early *Homo* has come from Koobi Fora in Kenya. I have been involved with *H. habilis* for all but two of its 50 years, starting in 1966, when I analysed the ankle bone excavated alongside Johnny's Child. Far from being like that of modern humans, the bone is a much better match for an australopithec. Other features of *H. habilis* have also turned out to be less like those of modern humans than Louis and his team suggested.

In the mid-1970s, Louis and Mary's second son, Richard, offered me the challenge of making sense of the early *Homo* skulls, crania and jaws from Koobi Fora. It was a lonely task involving 15 years poring over australopithecids and *H. erectus* fossils in museums around the world. It was tempting to focus on the better-preserved specimens, but more often than not it was a skull fragment here or a broken tooth there that provided the key clues to making sense of the whole collection.

Variation in the Koobi Fora fossils was not

so easily shoe-horned into a single species as those from Olduvai³. I concluded that there were two distinct types of face within early *Homo*⁴, and so in 1992, I suggested that a second early *Homo* species, *Homo rudolfensis*, should be recognized⁵. Two decades later, a team led by palaeontologist Meave Leakey (Richard's wife) confirmed⁶ the 'two-taxon' hypothesis I had proposed, using a face and two lower jaws found at Koobi Fora. But they — correctly, I believe — refuted my suggestion about which jaws went with which faces. As ever in palaeontology, new fossils test and refine old ideas.

DRAWING THE LINE

In 1999, British anthropologist Mark Collard and I looked⁷ afresh at the boundary between *Homo* and more-primitive hominins by focusing on features that hint at body size, posture, locomotion, diet and life history. For example, how long is the upper limb compared with the lower, or the forearm compared with the upper arm? Do molar teeth erupt early, as in apes, or form slowly and dawdle in the jaw, as in modern humans? All are attributes that help to reveal how an animal makes its living and allocates its energy.

Although *H. habilis* is generally larger than *A. africanus*, its teeth and jaws have the same proportions. What little evidence there is about its body shape, hands and feet suggest that *H. habilis* would be a much better climber than undisputed human ancestors. So, if *H. habilis* is added to *Homo*, the genus has an incoherent mishmash of features. Others disagree, but I think you have to cherry-pick the data⁸ to come to any other conclusion. My sense is that handy man should belong to its own genus, neither australopithec nor human.

Beautifully preserved fossils from the Caucasus have now been added to the mix. Just last year, Georgian anthropologist

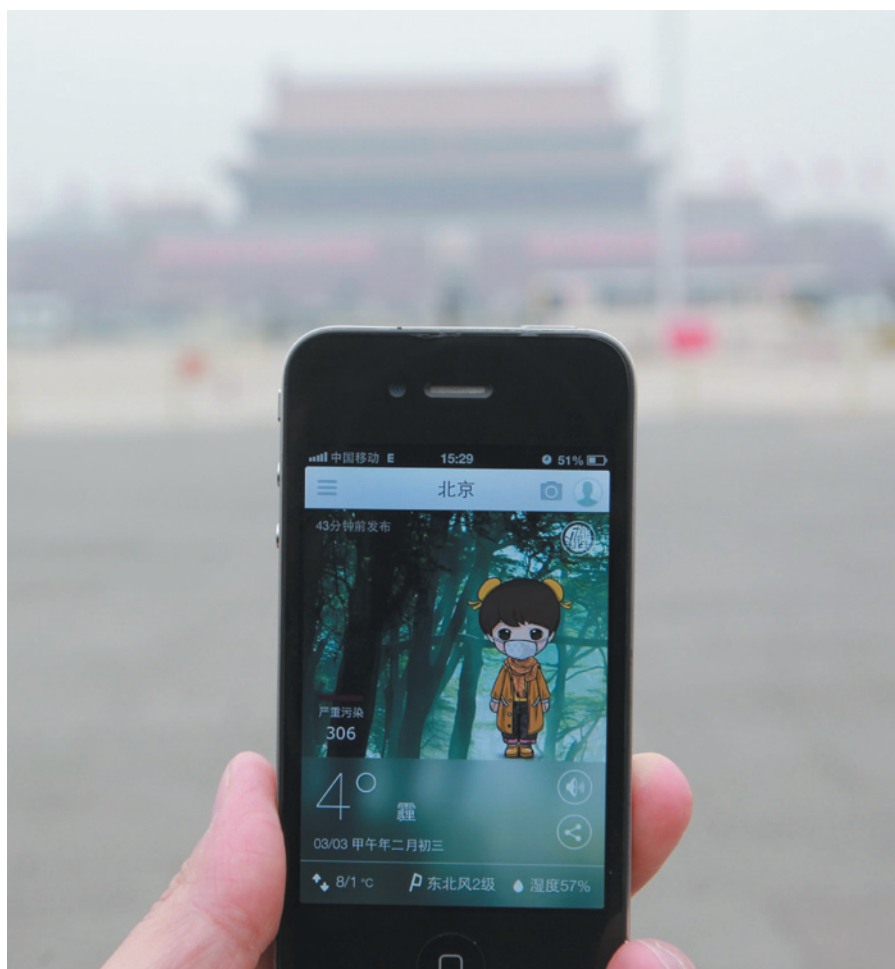
David Lordkipanidze and his colleagues reported⁹ their analysis of five hominin crania recovered from Dmanisi, a spectacular site on a promontory between two rivers in southern Georgia. They concluded that the range of shapes among these skulls equals or exceeds the variation across *H. habilis*, *H. rudolfensis* and *H. erectus*, and on that basis proposed that all *H. habilis*-like fossils be reassigned to *H. erectus*, subsuming three species into one.

Even if you accept that their methods of data capture are sound — which I do not — I question their conclusions. Their method fails to distinguish between a distinctive and large-brained Neanderthal cranium and one of the small-brained Dmanisi skulls, specimens that are separated by close to two million years of evolutionary history. They also take the overall shape of the head to be the arbiter of early hominin taxonomy, yet what sets *H. habilis* and *H. erectus* apart are many finer details, such as the size and shape of the inner ear, features of the hands and feet, the strength of long bones and life history. It is equally plausible that the Dmanisi fossils sample a hominin taxon that exhibits a hitherto unknown combination of primitive (for example, a small brain) and derived morphology (for example, brow ridges).

The ongoing debate about the origins of our genus is part of *H. habilis*'s legacy. In my view, the species is too unlike *H. erectus* to be its immediate ancestor, so a simple, linear model explaining this stage of human evolution is looking less and less likely. Our ancestors probably evolved in Africa, but the birthplace of our genus could be far from the Great Rift Valley, where most of the fossil evidence has been found. The Leakeys' iconic discoveries at Olduvai Gorge should remind us of how much we don't know, rather than how much we do. ■

Bernard Wood is a palaeoanthropologist at George Washington University in Washington DC, USA.
e-mail: bwood@gwu.edu

1. Leakey, L. S. B., Tobias, P. V. & Napier, J. R. *Nature* **202**, 7–9 (1964).
2. Le Gros Clark, W. E. *The Fossil Evidence for Human Evolution* (Univ. Chicago Press, 1955).
3. Tobias, P. V. *Olduvai Gorge: The Skulls, Endocasts and Teeth of Homo habilis* (Cambridge Univ. Press, 1991).
4. Wood, B. A. *Koobi Fora Research Project, Volume 4: Hominid Cranial Remains* (Clarendon Press, 1991).
5. Wood, B. A. *Nature* **355**, 783–790 (1992).
6. Leakey, M. G. et al. *Nature* **488**, 201–204 (2012).
7. Wood, B. A. & Collard, M. *Science* **284**, 65–71 (1999).
8. Holliday, T. W. *Curr. Anthropol.* **53**, S330–S345 (2012).
9. Lordkipanidze, D. et al. *Science* **342**, 326–331 (2013).



Apps to measure air quality proliferated in China following controversies with government statistics.

Mobilize citizens to track sustainability

Businesses and the public can keep watch when governments fail to provide environmental data, say **Angel Hsu** and colleagues.

United Nations negotiators are meeting in New York this week to shape up the Sustainable Development Goals (SDGs) that will replace the Millennium Development Goals (MDGs) after 2015. The scope of the SDGs — from providing universal access to energy and water to ending poverty by 2030 — is being well articulated. But there has been little discussion about how countries will monitor that progress.

The variety of global environmental information that will be needed raises daunting challenges. Official data sets are not up to the task. We have found problems with government-reported sources in nearly every global data set that we have used in 15 years of constructing the Environmental Performance Index (EPI) — a biennial ranking of how well countries are implementing policies ▶

► to address pressing environmental concerns (see epi.yale.edu).

Government investments in environmental monitoring, data collection and reporting are patchy, and are influenced by limited budgets and political motivations. Governments are notorious for underreporting fish catches¹, for example, and have been criticized for using capricious definitions, such as of what constitutes 'forest'². The global data sets that do exist are often incomplete, erratic or untrustworthy. Conspicuous reporting gaps compromise our understanding of most environmental problems, from toxic chemical exposures, global recycling rates and wetlands loss to freshwater quality, species loss and vulnerability to climate change.

As a result, the data required to track progress towards SDG targets cannot come solely from governments or intergovernmental organizations. UN negotiators must think more creatively about how to measure progress. We argue for channels by which citizen scientists, independent watchdogs, private-sector companies and third-party organizations can contribute data towards monitoring SDG progress and make governments more accountable. Without such independent monitoring, the extent of environmental challenges will not be captured, and SDG-related policies and management decisions risk being ad hoc.

SENSING EARTH

Because a pixel in one country is measured in the same way as one in another, satellite data will be invaluable in establishing baselines and benchmarks for the SDGs. Global forest loss over the past decade has been tracked in more than 650,000 images from NASA's Landsat programme by researchers at the University of Maryland in College Park and by the Google Earth Engine platform, for example³. To estimate how much water countries have underground, NASA and the German Aerospace Center's orbiting Gravity Recovery and Climate Experiment are following changing aquifer levels⁴.

Yet space imagery is rarely used in the public-policy sphere beyond a few applications, such as land-use planning. Computing power and scale are two reasons. The global forest-loss calculations took the equivalent of 1 million processing hours on 10,000 computers to process 20 terapixels of data — beyond the reach of most national statistical agencies. Many satellite-derived data sets are too coarse in resolution for local decision-makers to act on. Global climate data cannot help a mayor to understand how a city will be affected by rising temperatures.

And politics gets in the way. In our experience, many governments question satellite data when comparisons between their countries and others are unfavourable or reveal weaknesses in their



Students in Hong Kong test local water quality as part of the World Water Monitoring Challenge.

statistical reporting systems. Other global environmental measurements such as ocean acidity, which are only beginning to be tracked consistently, will also be susceptible to political pressures.

In January, we published an indicator based on global satellite-derived estimates of fine particulate air pollution in the 2014 EPI (see go.nature.com/tftogi). The Indian government, which like many developing countries does not regularly release such data, was shocked. Government officials in New Delhi were quick to refute the suggestion that their air quality might be as bad as Beijing's, igniting a public debate.

The misreporting of environmental data by governments is common. The New Zealand government, for instance, which touts a "100% pure" slogan to burnish its ecotourism reputation, was revealed in 2007 to have altered some unfavourable conclusions of a State of the Environment report. The country has not released a comprehensive State of the Environment report since. After demands from citizens and debates in parliament, legislation was introduced last August requiring that bodies independent from the government report on environmental conditions every six months and a compile synthesis report every three years.

ENGAGING CITIZENS

Enter citizens. They are increasingly contributing environmental measurements and geographical information, through social media, crowdsourcing and open-source

databases such as OpenStreetMap, a free and editable map of the world. The World Water Monitoring Challenge, for example, encourages people to record their local water quality and share results (see www.worldwatermonitoringday.org). In 2012, participants made some 250,000 visits to sites in 66 countries, recording parameters such as water temperature, pH and dissolved oxygen. Citizens even monitor plankton abundance in oceans (see www.playingwithdata.com) and donate spare computer time to run climate simulations (see www.climateprediction.net).

User-generated sources can gather more data than any government agency could manage. Every day, people upload roughly 100 terabytes of data to Facebook; send 294 billion e-mails; and write 230 million tweets. Mobile-phone users send and receive 1.3 exabytes of data and each household consumes 375 megabytes of data (see go.nature.com/fhzuqr).

Watches, tablets or phones equipped with sensors could allow millions of citizens to monitor pollutants. Prototypes range from hand-held devices that can measure air and water quality (such as AirBot and WaterBot, developed by researchers at Carnegie Mellon University in Pittsburgh, Pennsylvania); an electronic 'nose' that can detect toxic substances in the air; and sensors that can determine nitrate levels in food. Such gadgets now need to be made cheaper and more readily available (the WaterBot is priced at US\$99).

Citizens can also track progress using smartphone apps. Since 2013, the Water

Reporter app has allowed people in the Chesapeake Bay area in Maryland and Virginia to report local water pollution and other problems to local managers. SeeClickFix has since 2008 allowed city residents to report and track civic issues, from broken streetlights and fire hydrants to street crime. The app has addressed more than 800,000 requests in 170 cities and towns worldwide, including many in the United States, as well as Dublin and Buenos Aires, together representing 25 million people.

Chinese citizens have already seen the power of public data collection. In 2011, statistics from a monitor atop the US embassy in Beijing revealed that the air quality was much worse than government statistics claimed. The discrepancies led members of the public to monitor their own air quality, using backpacks, kites and smartphone apps, and a website charting the live results emerged.

Yet user-generated and crowdsourced data have not been discussed for the SDGs (although the Rio+20 United Nations Conference on Sustainable Development piloted an online voting system for delegates).

THIRD PARTIES

Businesses, too, might be better poised than governments to collect environmental data. Coca-Cola, the beverage company, operates in almost every country. It requires more than 9 litres of water to generate \$1 of revenue, so relies on accurate knowledge of water resources. Since 2004, the company has invested more than \$1.5 million in recording and assessing surface and groundwater levels, stresses and drought severity.

In 2011, the company teamed up with the World Resources Institute, an environmental think tank in Washington DC, to make its proprietary data publicly available through a web platform called Aqueduct. The company is keen to address criticisms of overextraction in water-stressed countries such as India, one of Coca-Cola's biggest growth markets. It also hopes to galvanize other businesses to evaluate their water impacts and to encourage government leaders in high-risk areas to manage water resources more efficiently, equitably and sustainably.

Third-party organizations can also validate data. The researchers in the Sea Around Us Project at the University of British Columbia in Vancouver, Canada, for instance, regularly reconstruct fish-catch data. They have noted that the UN's Food and Agriculture Organization underestimates the percentage of over-exploited and collapsed fish stocks, owing to its use of variable-quality government-reported data and its omission of data from other sources⁵. The resulting biased view of the status of the world's fisheries

could have disastrous consequences for global aquaculture and ocean health.

In India, the Centre for Science and Environment, a non-profit research organization based in New Delhi, has added its independent analysis to the capital's air-quality debate. It remains to be seen whether the Indian public will follow China's example and pressure the government for better air-pollution monitoring, or start monitoring pollution for themselves.

The full potential for private-sector and non-government engagement has not been explored in the SDG negotiations. The most recent progress report states that "business should be part of the solution", but only by encouraging "greater private sector uptake of sustainability reporting". A 2013 survey by financial services firm KPMG shows that 71% of companies worldwide are already doing this⁶. The crucial question is how companies can be incentivized to share data.

NEXT STEPS

So, what now? First, SDG negotiators should clarify in the next few months the pivotal questions that better data can help to answer. What knowledge is required for countries to transition to a green economy, for example? What factors define well-being in a society? Then

"If Coca-Cola collects the best global water data, then why not use them?"

they should consider how those data might be crowdsourced and contemplate incentives for participation. A fund could be created, for example, through contributions from countries, private foundations and companies or through crowdfunding, that would reward the individuals or institutions most capable of collecting needed data.

The UN should consider new forms of collaborations. If Coca-Cola collects the best global water data, then why not use them to measure progress towards a global water SDG? If Google is best able to process vast amounts of satellite data, why not work with it and other scientists to develop algorithms to assess pressing environmental issues?

To replicate the data-driven approaches to city management used by former New York City mayor Michael Bloomberg and Maryland governor Martin O'Malley, the UN should provide an online platform for cities to share relevant data on public safety, disaster preparedness and health. Individual innovators who invest in a low-cost technology to engage citizen scientists in data collection for an SDG should be funded to share devices globally.

Second, if citizens are to contribute data to the SDGs, protocols and guidelines must be established to protect individual rights

and privacy. Individuals should know how their data will be used and be assured that their privacy is maintained. A starting point for the international harmonization of privacy laws is the Organisation for Economic Co-operation and Development's Guidelines on the Protection of Privacy and Trans-border Flows of Personal Data.

In places where information and communication technologies are still emerging, more can be done to equip people with affordable tools to participate equally in the data revolution. The SDG process could foster technology-transfer or funding mechanisms to provide citizens in developing countries with free or cheap personal environmental monitoring devices or community-based systems.

Last, negotiators must find ways to incentivize participation. Corporate or private sponsorship of new data streams (which might not sit well with some audiences who fear commercialization of the SDGs, or worse yet, could have competing private interests that may bias data) could support innovative sustainability-minded companies or individuals to share data. A transparent, centralized online 'dashboard' would make it easy for citizens, businesses and third-party institutions to contribute and share data. This could be administered by the UN Environment Programme, which countries at the Rio+20 Earth summit pledged to bolster.

UN negotiators are running out of time to get the SDGs right. It is important to set appropriate targets for promoting sustainable development. Meeting those targets will depend on how well we can track progress, using the best data available, collected by the people and organizations best placed to do so. ■

Angel Hsu is director of the Environmental Performance Measurement Program at the Yale Center for Environmental Law and Policy, Yale University, New Haven, Connecticut, USA. **Omar Malik** and **Laura Johnson** are environmental performance analysts at the Yale Center for Environmental Law and Policy, Yale University, New Haven, Connecticut, USA. **Daniel C. Esty** is professor of environmental law and policy at Yale University, New Haven, Connecticut, USA. e-mail: angel.hsu@yale.edu

1. Pauly, D. & Froese, R. *Mar. Policy* **36**, 746–752 (2012).
2. Friess, D. A. & Webb, E. L. *Environ. Conserv.* **38**, 1–5 (2011).
3. Hansen, M. C., Stehman, S. V. & Potapov, P. V. *Proc. Natl Acad. Sci. USA* **107**, 8650–8655 (2010).
4. Rodell, M. & Famiglietti, J. S. *J. Hydrol.* **263**, 245–256 (2002).
5. Froese, R., Zeller, D., Kleisner, K. & Pauly, D. *Mar. Biol.* **159**, 1283–1292 (2012).
6. KPMG. *The KPMG Survey of Corporate Sustainability Reporting 2013* (KPMG, 2013).



ILLUSTRATION BY PHIL DISLEY

Free Indian science

As elections begin in India, **Mathai Joseph** and **Andrew Robinson** call for an end to the stultifying bureaucracy that has held back the nation's science for decades.

India's general elections this month and next could be among the most important since it gained independence in 1947. After ten years of a largely indecisive and an often scandal-ridden coalition government, there are strident demands for better governance, economic reform, the promotion of manufacturing and improvements in agriculture, health care and environmental management.

Sadly, science and its administration, once seen as central to Indian development, are not currently on the agenda, despite some trenchant critiques from scientists and science policy-makers^{1,2}. Repeated government promises to increase the expenditure on research and development (R&D) to 2% of India's gross domestic product have not been kept. R&D spend remains at about 0.9% of GDP — compared with 1.12% in Russia³ (down from 1.25% in 2009), 1.25% in Brazil and 1.84% in China² (see 'Brick benchmarking').

That said, the stagnation afflicting

Indian science is as much structural as it is financial. Before the machinery of government took over and mismanaged research in the mid-twentieth century, several foundational scientific discoveries were made in India. Between about 1900 and 1930, Jagadish Chandra Bose made innovations in wireless signalling (borrowed by Italian electrical engineer Guglielmo Marconi); Meghnad Saha developed an ionization formula for hot gases that has a central role in stellar astrophysics; Satyendra Nath Bose's theoretical work in quantum statistics led to Bose–Einstein statistics; Chandrasekhara Venkata Raman did Nobel-prizewinning work on light scattering; and in mathematics, Srinivasa Ramanujan was equally pioneering.

But since 1947, there has not been a single Nobel-prizewinning scientific or technological discovery, despite India's successes in space, radio astronomy, biology and pharmaceuticals and the worldwide reputation of its US\$100-billion information technology (IT)

industry. Three other Indian-born scientists have won a Nobel prize — biochemist Har Gobind Khorana (in 1968), astrophysicist Subrahmanyan Chandrasekhar (in 1983) and molecular biologist Venkatraman Ramakrishnan (in 2009) — but for work done entirely outside India. No mathematician from India has won the Fields Medal. And Indian institutes and universities do not feature in the world's top 200 higher-education institutions (see go.nature.com/bc69uq).

The basic problem is that Indian science has for too long been hamstrung by a bureaucratic mentality that values administrative power over scientific achievement. And, to preserve local control, research is still done mostly by small teams working in isolation rather than through collaboration — a key generator of impact⁴.

More than two decades ago, the threat of imminent national bankruptcy forced India's government to liberate its economy from the notorious 'licence-permit raj', which had strait-jacketed commerce and industry since

1947. What will it take in 2014 to reinvigorate India's decrepit scientific empires, trapped for decades in a similarly rigid bureaucracy?

DEEP-ROOTED PROBLEM

The problem has a long history. The Council of Scientific and Industrial Research (CSIR) was formed in 1942, before independence, to establish five national laboratories aimed at converting research discoveries into industrial applications. It was soon widely derided. Raman, referring to the first director-general of the CSIR, chemist Shanti Swarup Bhatnagar, said: "Bhatnagar built the National Laboratories to bury scientific instruments"⁵. The situation today is no better. A former CSIR director-general, chemical engineer Raghunath Anant Mashelkar, remarked in 2013: "India can't remain a nation of imitators."

In 1954, India's Department of Atomic Energy (DAE) was created using a different model, later replicated for other scientific departments, such as those for space, science and technology, electronics, biotechnology and ocean development. Its first head, nuclear physicist Homi J. Bhabha, was made a secretary to the government, on a par with top administrators in home affairs, finance and defence. This gave atomic energy official credibility, but placed it in a bureaucracy that was not designed to foster innovation.

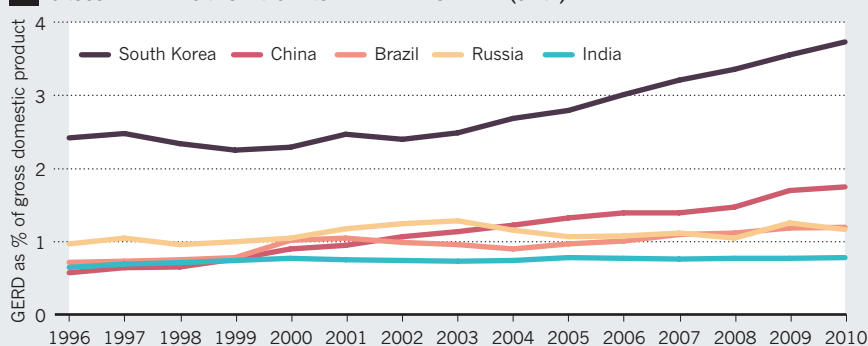
Gradually, the DAE's independence was ground down and its scientists and technologists slotted into administrative grades in which they could progress no faster than their non-scientific peers. Research achievement offered few rewards, other than a patriotic pat on the back. The other scientific departments quickly went down the same route. Scientists began to measure success by their administrative position and left research to juniors, who saw what they had to do to move up the hierarchy. If good science was done along the way, it was incidental. Today, although India ranks tenth in the world for output of scientific papers, it ranks 166th for average citations per paper (see go.nature.com/xl3ldg). Almost 20% of patents filed at the World Intellectual Property Organization in 2010 were from China, with just 1.9% from India (below Russia's 2.1% but above Brazil's 1.1%)⁶.

Nearly 60% of India's science budget⁷ is now spent on the CSIR, scientific departments and the Defence Research and Development Organisation (DRDO) — an enormous and impenetrable empire set up in 1958. None of these national institutions has stimulated scientific excellence. Indian scientists do outstanding work, but not in India. The latest examples of this long-familiar situation include the award of the 2014 Marconi Prize in the United States to engineer Arogyaswami Joseph Paulraj of Stanford University in California, who worked in the Indian navy and at the Centre for

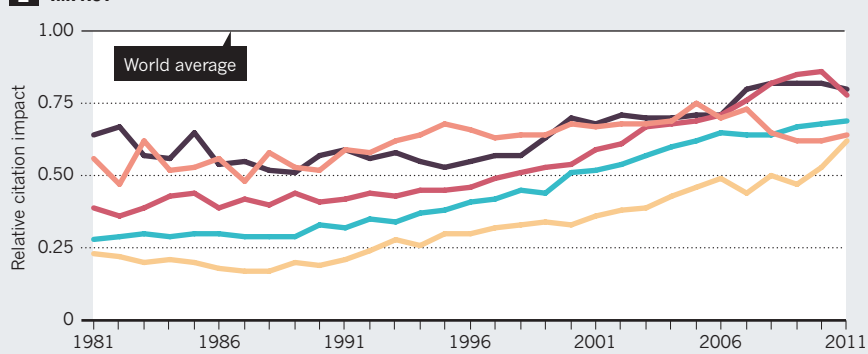
BRICK BENCHMARKING

Of the emerging economic powers, India spends the least on research and development. On citations, it tails China, South Korea and until recently, Brazil.

1 GROSS EXPENDITURE ON RESEARCH AND DEVELOPMENT (GERD)



2 IMPACT



Development of Telematics for some years before emigrating in the early 1990s, and the appointment in February of Indian-born Satya Nadella as chief executive of Microsoft.

The problems at the national level are mirrored in institutions. First, scientists are promoted on the basis of years of service, rather than achievement, and once at the top they stay until retirement age; long after, in some cases. Even at the prestigious Tata Institute of Fundamental Research (TIFR) in Mumbai, which is less rule-bound than many other institutions, research groups are almost invariably headed by those who have been there the longest.

Second, although research in the leading institutions is well funded — with more money available than requested in credible grant applications, a striking contrast to the situation in many nations — the funding is subject to unsuitable restrictions applicable to the entire government bureaucracy. These include limited foreign travel and no travel support for research students, ruling out regular participation in leading conferences and research gatherings.

Third, the movement of researchers from one institution to another is discouraged, because administrators prefer senior positions to be filled by internal promotion rather than lateral hiring.

One would expect respected bodies of scientists to question the government's

virtual abandonment of science. But none of India's science academies (such as the Indian National Science Academy and the Indian Academy of Sciences) has taken any action — even on the widely reported cases of plagiarism by their fellows⁷.

FOUR STEPS TOWARDS CHANGE

Indian science needs public funding, but not government control. In many countries, the promotion of science is devolved to agencies outside the main government structures, such as the United Kingdom's Engineering and Physical Sciences Research Council, the European Research Council, the US National Science Foundation and Singapore's Agency for Science, Technology and Research.

The first step towards reinvigorating Indian science must be to create an empowered funding agency, staffed by working scientists, some of whom could be non-resident Indians. A possible model is the European Research Council, which deals with a complex of national governments no less formidable than India's 29 state governments, yet manages to focus on supporting research excellence. The crucial requirement is obviously that an Indian scientific research council be permitted to set its own criteria for the evaluation of research proposals, independent of direct government control, and disburse government funds accordingly.

A second step must be to ensure

planned rotation of institutional roles and responsibilities. This occurs in most university departments in the Western world — typically, every four or five years for the chair of a UK university department. Governing bodies should limit the tenure of the heads of scientific institutions and groups to, say, five years, after which they would be expected to return to active research. This change would work best by choosing heads young enough to have future research careers. Bhabha was 35 years old when he was appointed director of the TIFR in 1944; his example has not been repeated.

Third, the formation of trans-institutional groups that can undertake coordinated work in a few well-chosen areas should be encouraged at the funding stage. This contrasts with the existing 'national missions' of the government. The \$160-million Nano Mission (launched in 2007) has funded more than 150 individual projects, 11 centres of excellence and 6 industry-linked projects — but has required no collaboration between institutions.

For building competence and achieving results, it would have been much more effective to encourage collaborative efforts across institutions in, for instance, medical applications, solar and fuel cells, and water purification. Such collaboration has been achieved successfully, for example, in the European Strategic Program on Research in Information Technology, the projects of which span several countries and agencies.

Fourth, how to spend that 2% of GDP when it finally materializes? Leading institutions such as the Indian Institutes of Technology and many others are already well provided for, by any standards⁸. New research money should be spent on regenerating the scores of poorly provided university laboratories that lack the funds to procure and maintain modern scientific equipment; they currently receive only around 10% of the R&D budget but are expected to produce most of the country's PhDs².

LESSONS FROM COMPUTING

Indian scientists working in conventional disciplines will be loath to admit it, but there is now a model of technological success in India — in the growth of the IT industry. The creation and export of software to the developed nations grew, even during the licence-permit raj of the 1970s and 80s, because software did not fall into any government category⁹. Its commercial success was driven primarily by young engineers in their 20s working in a competitive environment unfettered by government regulations.

Ironically, academic computer science

in India barely benefited from this boom until recently, when a few enterprising IT companies such as Tata Consultancy Services, Microsoft and Infosys instituted well-planned funding to pay attractive stipends to young computer scientists taking up research careers.

One lesson from India's IT industry is that it is essential to draw the private sector into major research programmes. Industry at present contributes about 30% of India's total spend on R&D², most of it devoted to improving productivity and reducing cost and energy consumption, rather than product development. It is essentially shut out of basic research because of government rules that prohibit, or severely inhibit, public-private collaboration. Although the massive public-sector defence industry relies mostly on purchasing foreign-made weapons and little on the laboratories of the DRDO, it is still unwilling to partner with Indian companies to grow competence and capability.

Another lesson is that science could attract talented young people if it provided them with a more exciting work environment and a career path that rewards achievement. A scientific career has the potential to be at least as challenging and stimulating as one in IT, even if not as financially rewarding.

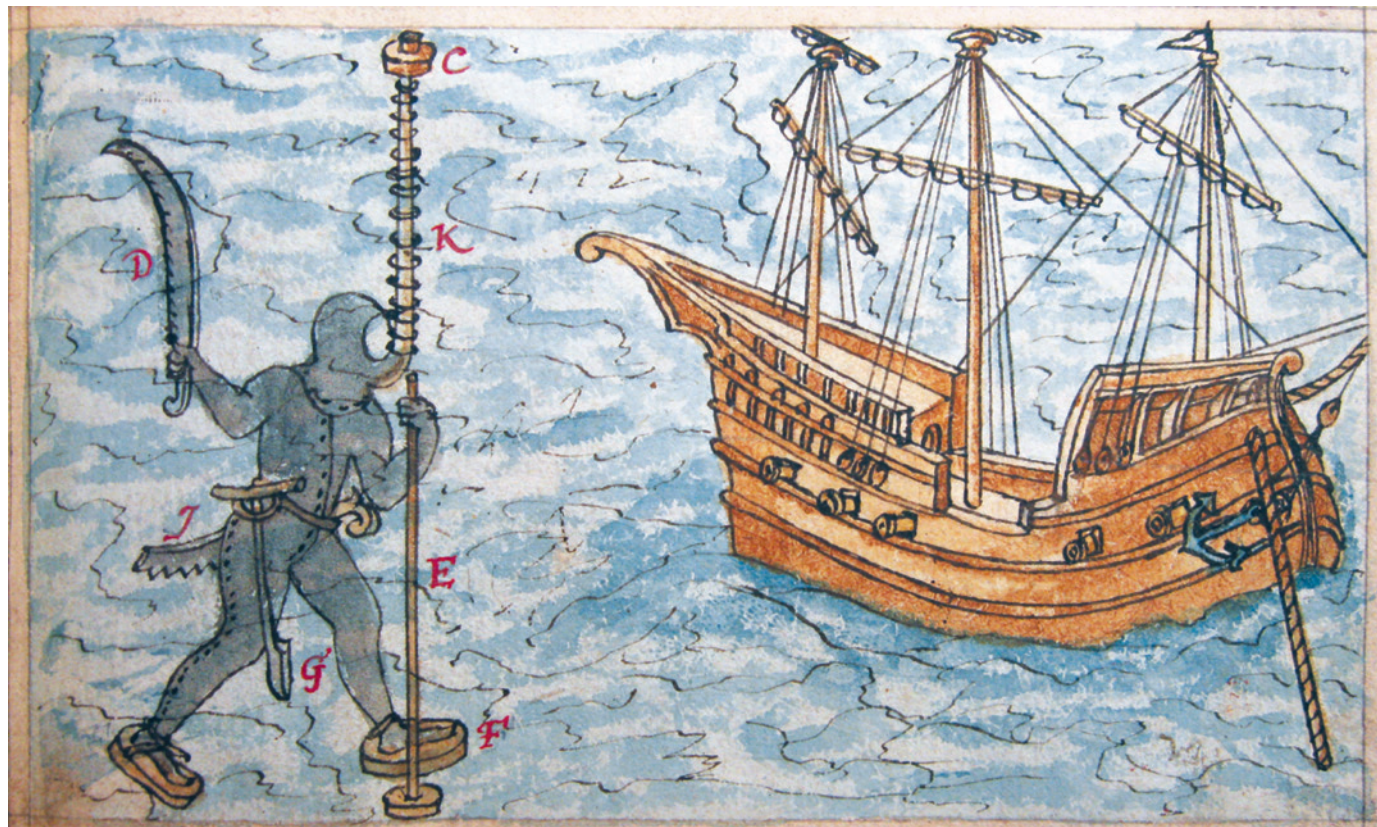
The Indian pioneers of the early twentieth century, such as Raman, made their theoretical and experimental breakthroughs with almost no government support; their research suffered from government apathy but not bureaucratic interference. The strong urge for discovery that drove them could return — if there were greater rewards for innovation, fewer for administration and longevity. ■

Mathai Joseph is a computer scientist and independent consultant. He was senior research scientist at the Tata Institute of Fundamental Research, Mumbai, India, and professor at the University of Warwick, UK. He was also head of research at Tata Consultancy Services from 1999 to 2007.

Andrew Robinson is the author of India: A Short History and Genius: A Very Short Introduction.

e-mail: andrew.robinson33@virgin.net

- Desiraju, G. *Nature* **484**, 159–160 (2012).
- Krishna, V. V. 'Paralysis in science policies' *The Hindu* (7 February 2014); available at <http://go.nature.com/rdqdf0>.
- World Bank *World Development Indicators: Research and Development Expenditure (% of GDP)*; available at <http://go.nature.com/x3ohil>.
- Adams, J. *Nature* **490**, 335–336 (2012).
- Parameswaran, U. C. V. *Raman: A Biography* 224 (Penguin, 2011).
- World Intellectual Property Organization *2012 WIPO IP Facts and Figures* (WIPO, 2012); available at: <http://go.nature.com/qjwanp>.
- Jayaraman, K. S. *Nature* <http://dx.doi.org/10.1038/nature.2012.10102> (2012).
- Varma, G. B. S. N. P. *Science* <http://dx.doi.org/10.1126/science.caredit.a1400003> (2014).
- Joseph, M. *Digital Republic: India's Rise to IT Power* (Power, 2013).



A prototype wetsuit — one of several innovations presented in the 1590s manuscript MS Osborn a8.

TUDOR TECHNOLOGY

Shakespeare and science

To mark the 450th anniversary of the bard's birth, **Jennifer Rampling** probes how mathematics and technology shaped his era.

Like the plots of William Shakespeare's plays, from *The Merchant of Venice* to *Hamlet* and *Othello*, Tudor science took its inspiration from abroad. However, its dramatis personae include few easy analogues of continental stars such as Nicolaus Copernicus, Johannes Kepler, Paracelsus or Galileo. England and Wales undoubtedly had scholars of international calibre, including the mathematician-astronomers John Dee, Thomas Digges and Thomas Harriot. Yet the expansion of its scientific horizons was driven as much by artisanal and mercantile interests as by university learning or royal patronage. Detailed globes, astronomical instruments and translations of French and Spanish navigational texts offered new tools for mastering both England's coastal waters and the ocean voyages required for intercontinental trade and exploration.

Shakespeare's "scepter'd isle" was a late-comer to Europe's scientific renaissance. The upheavals of the English Reformation

diverted energies that might have otherwise been channelled into exploration. By the time of Elizabeth I's accession in 1558 (six years before the playwright's birth on 23 April 1564), her Spanish and Portuguese counterparts had amassed vast New World territories — and with them, knowledge of new plant and animal species, peoples, geographies and commodities. England's rotting fleet and dearth of native pilots and navigational know-how offered dispiriting prospects for the expansion of either territory or natural knowledge.

A quick scene change to almost 40 years later, towards the end of Elizabeth's reign, and we find the port of London booming and England a maritime power, defending its shores from Spanish fleets and establishing fledgling colonies in the Americas. Goods, people and ideas poured in from continental Europe and beyond, bringing new experiences and expertise. London's presses added to the flow of information. Scientific

best-sellers included almanacs, collections of alchemical and household 'secrets' and vernacular mathematical textbooks such as Robert Recorde's oft-reprinted *Arithmetic, or The Grounde of Arts* (1543).

Enter Shakespeare. Born in Stratford-upon-Avon, by 1592 he was in London as both player and playwright. From temporary stages in inn yards to purpose-built theatres, the capital offered increasing scope for companies of players. There, a growing population of 'mechanical' artisans, instrument-makers, engineers, printers and medical practitioners helped to constitute one of the most sophisticated play-going audiences in history, eager for new material and swift to detect topical allusions. Whatever the causes and effects of the English transformation, we might expect to catch some hint of it on the stage.

For both playwrights and audiences, the connection between global ambition and artisanal know-how would have been hard to miss: new technologies simultaneously ►



Elizabeth I rests her hand on a globe — an instance of technology advertising naval prowess.

► enabled expansion and advertised it. In 1592, the Lambeth-based mathematician and instrument-maker Emery Molyneux presented Elizabeth I with intricate, 65-centimetre globes that tracked the voyages of the English explorers Francis Drake and Thomas Cavendish. The globes' fame helped to promote the use of smaller versions as navigation aids, and even reached the stage: playwright Thomas Dekker referenced "Molyneux his globe". In 1599, Shakespeare's company, the Lord Chamberlain's Men, named their new theatre the Globe — a fitting symbol of Elizabethan aspirations. The unexpected richness of New World discoveries, and the challenges they posed to ancient authorities, forced Europeans to expand their world views even as their voyages put a girdle round the Earth.

Shakespeare's famously sparse biography offers few clues to his own views on the sciences. Young Will may have witnessed the blazing Stella Nova: the supernova spotted by Danish astronomer Tycho Brahe and others in 1572. Unlike comets, the new star had no apparent parallax or proper motion, suggesting that it lay at a great distance from Earth — a novelty where none should have existed, according to the generally accepted Aristotelian model of the cosmos. For Digges, the star offered a potential confirmation of the Copernican system, and an opportunity to advance his reputation through timely publication of his own work on stellar parallax.

Not all observers read the star in that way. For astrologically literate Londoners, stellar novelties might also presage warfare and dynastic change. Shakespeare well

understood the analogy between heavenly and political stability, and the anxieties triggered by inauspicious stars, comets and sunspots (the "disasters in the sun" mentioned by Horatio in *Hamlet*). The playwright's dramatic use of celestial portents does not stretch to promoting a Copernican model, although we may detect a glancing reference to heliocentrism in Hamlet's love letter to Ophelia — "Doubt thou the stars are fire; Doubt that the sun doth move ... But never doubt I love". Only after Shakespeare's death and Galileo's energetic interventions would Copernican ideas take centre stage.

Few Elizabethan scholars pursued astronomy on the scale of Brahe or Galileo, both recipients of princely patronage. (An exception was Harriot — the protégé of Walter Raleigh and Henry Percy, Earl of Northumberland — who mapped the Moon by telescope before Galileo.) Elizabethan science, like Elizabethan dramaturgy, responded to funding priorities, and theoretical advances carried less weight with Elizabeth's government than pragmatic and potentially high-profit schemes to improve navigation, exploit mineral resources or alchemically transmute metals on an industrial scale.

Petitioners to the queen often sought monopoly grants for such projects. Others hoped for financial investment. Apart from utility and personal expertise, proposals

"Elizabethan science, like Elizabethan dramaturgy, responded to funding priorities."

drew on classical precedents and the latest knowledge from abroad. Grenades, printing presses and fireworks attached to the backs of cats (intended to "raise a tumult" in enemy camps) appear in one manuscript collection of Elizabethan inventions. The anonymous compiler (possibly the engineer and alchemist Ralph Rabbards) combined practical experience of Italian military campaigns with fanciful ideas for adapting ancient inventions — among them, an ingenious design for a wetsuit with adjustable snorkel, for covert attacks on enemy shipping.

The tension between learning and experience, speculation and utility, can also be detected in scholarly works. When John Dee famously surveyed the "Artes Mathematicall" in his preface to Henry Billingsley's 1570 English translation of Euclid, he stressed the value of mathematics in studying both natural phenomena and practical problems — from arranging artillery to compounding medicines. The utilitarian ethos helps to explain why Digges, for all his astronomical skill, served Elizabeth primarily by consulting on the construction of Dover Harbour. Dee, who coined the term "Brytish Empire", eventually left England to pursue his dream of a court-philosopher post abroad.

Only in *The Tempest* does Shakespeare evoke the kind of philosopher that Dee sought to embody: the magus Prospero, whose expertise in natural and occult philosophy grants him the upper hand over the unschooled Caliban and the shipwrecked royal party. Dee, who tried and failed to persuade Elizabeth to sponsor a 'research institute' under his direction, might well have envied Prospero the freedom to experiment without the constraints of patronage. On the whole, the "brave new world" of Tudor knowledge-making lay elsewhere — between the commercially minded Antonio, the eponymous merchant of Venice, and the "rude mechanicals" of *A Midsummer Night's Dream*.

This is the world of Shakespeare's plays, populated with ideas and technologies that his audience would have recognized from contemporary life: clocks, globes, compasses, the distorting 'perspective glass'. Even his famous evocation of England, a "precious stone set in the silver sea", suggests the technique of foiling to create brilliant reflective backings for jewels and mirrors. If Shakespeare declined to furnish later ages with a convenient narrative of Elizabethan science, he surely succeeded in holding up a mirror to the technological and commercial vibrancy of his own age. ■

Jennifer Rampling is a historian of medieval and early-modern science and medicine. An assistant professor of history at Princeton University in New Jersey, she specializes in the history of alchemy. e-mail: rampling@princeton.edu



James Lovelock in his lab — inventions, notebooks and other materials from which will go on show at the Science Museum in London from 9 April.

EARTH SYSTEMS

No place like home

The newest chapter in James Lovelock's Gaia saga holds out hope, finds **Tim Lenton**.

In *A Rough Ride to the Future*, James Lovelock argues that humans must survive difficult days ahead if their descendants are to become the future saviours of Gaia — the self-regulating Earth system that he conceived nearly 50 years ago. No longer just another species, as he argued in *The Ages of Gaia* (Oxford University Press, 1988), *Homo sapiens* is now “one of the truly important species of the Earth's history”, destined to help Gaia to survive in the face of an ineluctably brightening Sun. The inventor's audacious book is thus part science fact and part science fiction, homily and history lesson.

Lovelock argues that our information-based cultural evolution is now a million times faster than good old natural selection. He traces this remarkable transition back to Thomas Newcomen's invention of the steam engine in 1712. That coal-powered engine's capacity to perform sustained useful work exceeding 1 kilowatt is Lovelock's thermodynamic definition of the start of the Anthropocene — the planetary epoch defined by human transformation. Powered by fossil fuels, the ensuing positive feedbacks have propelled an exponential growth in material flows, human population and information processing.

But human innovation is just the latest in a long line of biological ‘innovations’ that have transformed Earth. For example, roughly 3 billion years ago, photosynthetic bacteria evolved the ability to split water, leading to the build-up of oxygen in Earth's atmosphere. Lovelock argues that this innovative capacity of life was crucial to the making of a world that supported human evolution, is crucial to our reshaping of the planet today, and will be essential for both our survival and Gaia's.

So, how does Lovelock rate our chances of making it through this century? Having heralded the apocalypse in *The Revenge of Gaia* (Allen Lane, 2006) and backed off a bit in *The Vanishing Face of Gaia* (Allen Lane, 2009), here he tries to shut the stable door after the horse (climate change) has bolted. He sees no prospect of us collectively reducing carbon dioxide emissions, and has nothing good to say here about geoengineering.



A Rough Ride to the Future: The Next Evolution of Gaia
JAMES LOVELOCK
Allen Lane: 2014.

That leaves adaptation as the only option.

Lovelock pictures most of us retreating to high-tech cities protected from the ravages of climate change by air conditioning — if we can afford it. He sees a glimmer of hope that a dwindling supply of cheap fossil-fuel energy may be slowing our exponential growth in material consumption. Meanwhile, to power those high-tech cities, he remains a consistent fan of nuclear power and even shows a soft spot for natural gas, while continuing to critique the weak and intermittent supply of wind energy. He fails to discuss solar power, perhaps because he sees an unlimited supply of renewable energy as a great danger to us and to the planet — we are, collectively, still too immature to handle it wisely.

Lovelock sees us as tribal carnivores, doomed to bumble around in the technological world we have created. But at long last he seems to be conceding that we might grow culturally to the point of becoming a constructive part of Gaia. This is the heart of the book — the vision of a teleological future Gaia, filled with foresight and purpose to survive, thanks to the consciousness bestowed on her by humanity. For example, we can consciously intervene to protect Gaia from future asteroid impacts. The irony here

is that teleology is exactly what Lovelock had to purge from his original Gaia hypothesis to satisfy his scientific critics. They may well squirm at the sight of “Gaia’s goal” returning.

The reason Gaia needs our help, Lovelock argues, is that the unconscious negative-feedback mechanism that has so far kept the planet cool is destined to fail. Up to now, the Sun’s brightening, which tends to warm Earth, has been counteracted by accelerated weathering of the continents, which has decreased CO₂ levels in the atmosphere. But before we started boosting it, the CO₂ concentration was getting close to the lower limit for photosynthesis. Hence there is a need for either some other cooling mechanism, or a more heat-tolerant intelligent life form.

Lovelock tentatively speculates on what such creations might look like. First he imagines us symbiotically merging with electronic components to form a world of cyborgs, using biological and artificial intelligence to sense and regulate the planet’s state. Eventually he sees wet organic life having to be replaced by dry electronic life that is better able to cope with an over-hot Sun. He even imagines such artificial life forms ‘evolving’ much faster than us, because information is transferred much faster by electronics than by neurons.

This is a popular science book. Yet it does reveal something about how the author’s remarkable mind works — making intuitive leaps between topics that most of us would not think to connect. One moment he is dwelling autobiographically on the cultural environment that fostered his own inventiveness, the next he is considering how invention accelerated the evolution of life. The leaps are so great that at times even Lovelock seems unsure where his argument is going. But this is strangely appropriate when the subject is the long-term future.

A Rough Ride to the Future is a more optimistic book than its two predecessors, although Lovelock still thinks that there will be a massive drop in human population — by means that remain unclear. What he seems to be doing, perhaps unwittingly, is creating a belief system to help us through the difficulties ahead. There is a sense here that Lovelock has been to the mountain top, looked over, and seen a promised land. Now almost 95, he may not get there with us, but he wishes humanity and Gaia an exciting future together. ■

Tim Lenton is professor of climate change and Earth system science at the University of Exeter, UK.
e-mail: t.m.lenton@exeter.ac.uk

PSYCHOLOGY

The clamorous mind

Susanne Ahmari reviews a personal and scientific journey through obsessive-compulsive disorder.

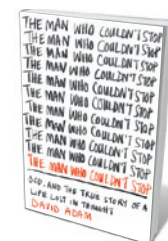
The adage that you can’t judge a book by its cover fails with David Adam’s *The Man Who Couldn’t Stop*. The title is scrawled repeatedly in urgent letters, thrusting us into the mindset of someone suffering a near-constant barrage of intrusive thoughts — the manifestation of obsessive-compulsive disorder (OCD) that is the book’s topic. In the pages that follow, Adam, editor of *Nature*’s Editorials section, leads us on an insider’s tour of the OCD brain, providing insight into the cultural and scientific evolution of how we view and treat a disorder that affects up to 3% of people worldwide.

Adam’s book is equal parts historical journey, scientific discussion and stark personal memoir of his battle with often shocking thoughts that come out of nowhere, the symptom of OCD that he describes as “a snowflake that fell from the summer sky”. A common example of such thoughts is the fear of harming someone else despite having no intent or desire to do so. For Adam, they took the form of fears that he had accidentally contracted HIV. Although his book pays homage to earlier works, including Judith Rapoport’s iconic *The Boy Who Couldn’t Stop Washing* (E. P. Dutton, 1989), it has a unique voice born of first-hand observations of Adam’s own changing mental processes, and his expertise as a science writer.

Although his historical anecdotes are informative and entertaining, Adam is at his best when linking the evolution of his illness with possible scientific explanations. Current theories suggest that OCD stems from abnormal communication in the brain between cortical areas involved in higher-order cognition — such as reward-based decision-making — and basal ganglia structures important for selecting actions. Human studies suggest that genetic and environmental factors contribute to this disruption. However, despite intensive efforts, it has been difficult to nail down specific genetic culprits, and the role of environmental factors, such as childhood infection with streptococcus, is unclear

and controversial.

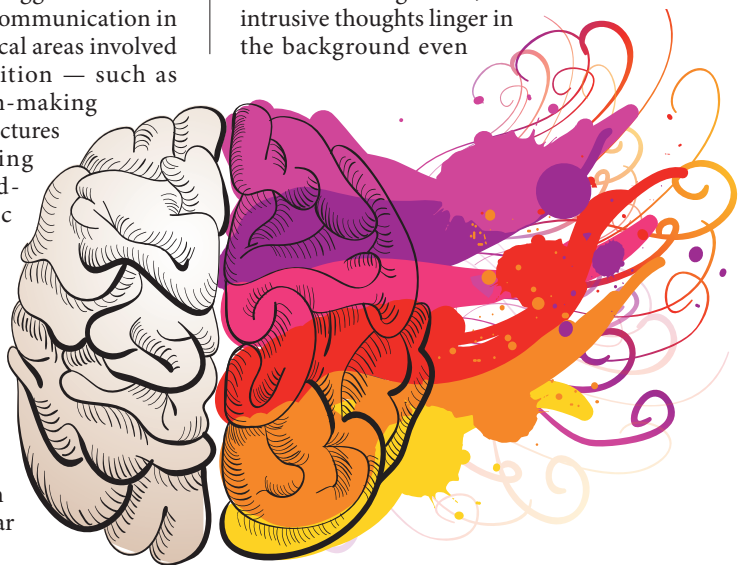
Accounts from sufferers may provide hints about how OCD evolves. In a particularly evocative story, Adam describes his attempts to combat the barrage of unwelcome thoughts about HIV by repeatedly calling the UK National AIDS hotline to check whether he was at risk. No matter how many times he called, disguising



The Man Who Couldn't Stop: OCD, and the True Story of a Life Lost in Thought
DAVID ADAM
Picador: 2014.

his voice to avoid recognition, he was only briefly reassured. He notes, “They would tell me that they had already given me an answer and that I needed to accept it. I didn’t want that. I wanted the hit.” This comparison between taking an addictive drug and seeking reassurance highlights important potential research avenues for neuroscientists trying to understand how inexorable cycles of obsessions and compulsions are initiated and maintained.

As Adam himself points out, the ending of the book is a little unsatisfying. There is no dramatic revelation of the root of his illness, no ‘eureka’ moment when scientists discover the cause of obsessions. His OCD gets significantly better with a lot of hard work, serotonin-reuptake inhibitors, exposure therapy with response prevention, and continuous vigilance, but intrusive thoughts linger in the background even



VENIMO/SHUTTERSTOCK

though he has received the best evidence-based OCD treatment. As he puts it, “it’s a bit like being a recovering alcoholic. You are always a certain number of days past your most recent obsessive-compulsive episode.”

Fortunately, misunderstanding about OCD, such as thinking that people with the condition are simply ‘neat-freaks’, is gradually decreasing. Adam makes an important contribution by increasing public awareness, and clarifying that OCD is not simply an exaggerated desire for hygiene and order, but rather a serious illness with many different symptoms caused by dysregulation of the brain’s anxiety and cognitive-control circuits. He also helps to defuse the stigma associated with thoughts that many patients find shameful.

Another source of optimism is the abundance of innovative technologies that are being used to identify treatment targets for compulsive behaviours. For example, my laboratory uses miniature microscopes to peer inside the rodent brain and observe how neural activity changes as behaviours such as repetitive grooming develop. Such

“OCD is a bit like being a recovering alcoholic. You are always a certain number of days past your most recent episode.”

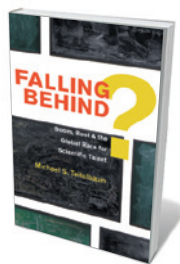
tools allow us to address long-standing questions, including how compulsive behaviours and anxiety are related.

Although the main readership of *The Man Who Couldn’t Stop* is likely to be patients, families and clinicians, this book is also a valuable read for neuroscientists like me, who are searching for clues about the origins of OCD. There are some factual errors, such as the description of serotonergic neurotransmission, but this is a quibble compared with the book’s contribution. People such as Adam hold in their brains some of the most valuable contributions to the fight against neuropsychiatric diseases: blow-by-blow accounts of how their obsessions, compulsions, thoughts and emotions are linked and evolve.

By partnering with patients, psychiatric researchers can extract testable hypotheses and recognize obsessions for what they are: not untouchable thoughts to be feared and locked up, but phenomena to be observed, understood and then let go like summer snowflakes melting in the sunlight. ■

Susanne Ahmari is director of the Translational OCD Laboratory at the University of Pittsburgh in Pennsylvania, where she uses neuroscience and clinical studies to seek brain abnormalities that lead to OCD.
e-mail: ahmarise@upmc.edu

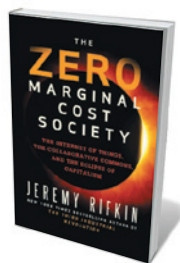
Books in brief



Falling Behind? Boom, Bust, and the Global Race for Scientific Talent

Michael S. Teitelbaum PRINCETON UNIVERSITY PRESS (2014)

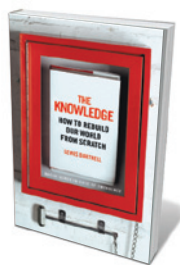
Is US science and engineering really plagued by workforce shortages? It is not that simple, argues demographer Michael S. Teitelbaum in this analysis of US science policy. The evidence reveals a complex dynamic: since the 1950s, alarmism over ‘falling behind’ has driven five rounds of destabilizing boom and bust. The modus operandi of research universities and funding agencies are also problematic, as research funding, not market demand, controls supplies of young scientists and engineers. Lucid and convincing.



The Zero Marginal Cost Society: The Internet of Things, the Collaborative Commons, and the Eclipse of Capitalism

Jeremy Rifkin PALGRAVE MACMILLAN (2014)

“The capitalist era is passing,” proclaims social theorist Jeremy Rifkin in this manifesto for the collaborative commons, a new economic paradigm. “Ubiquitous computing” is spawning an Internet of Things, a digitized global infrastructure that will, he argues, pare marginal costs of production and distribution down to near zero and topple big profits. The thesis is admirable, but the ebullience can feel somewhat relentless as Rifkin rolls out a future of sustainable abundance, massive open online courses and crowd-funding.



The Knowledge: How to Rebuild Our World from Scratch

Lewis Dartnell BODLEY HEAD (2014)

Astrobiologist Lewis Dartnell offers the ultimate do-it-yourself guide to ‘rebooting’ human civilization after an asteroid impact, pandemic or nuclear war. With scientific nous, Dartnell depicts probable environmental scenarios on a stricken Earth and offers putative survivors instruction in the technologies needed to craft a culture from the ground up. Whether learning the basics on how to sow and grow a field of barley, melt and cast salvaged aluminium, build a windmill or craft a basic stethoscope, many will thrill to this reminder of our species’ prodigious resilience.



How to Make a Human Being: A Body of Evidence

Christopher Potter FOURTH ESTATE (2014)

In this follow-up to his much-vaunted *You Are Here* (HarperCollins, 2009), Christopher Potter illuminates the human in all its manifestations — from single cell to creator of culture. Finely judged quotes from scientific and literary luminaries such as John Archibald Wheeler and Marcel Proust alternate with erudite musings on our compatibility with the cosmos (backed by clear expositions of theoretical physics), human biology, neurology, culture, morality and religion. The scattershot narrative somehow coalesces into a brilliant whole — and a compelling case for anti-reductionism.



Buildings Must Die: A Perverse View of Architecture

Stephen Cairns and Jane M. Jacobs THE MIT PRESS (2014)

From the passage of time to the wrecking ball, the forces limiting a building’s ‘lifespan’ are rife: all cities are potential ghost towns. Into this rubble-strewn terrain venture Stephen Cairns and Jane M. Jacobs, whose philosophical meditation on architectural death is enriched with materials science and thinking from the likes of architect Eero Saarinen. The evocative case studies range from the “ruin porn” of derelict Detroit in Michigan, to the weird morphology of Bangkok and its unfinished skyscrapers. **Barbara Kiser**

Correspondence

ArXiv screens spot fake papers

Unlike the computer-generated nonsense papers in some peer-reviewed subscription services (see *Nature* <http://doi.org/r3n>; 2014), the 500 or so preprints received daily by the automated repository arXiv are not pre-screened by humans. But sometimes automated assessment can be better than human diligence at enforcing standards.

The automated screens for outliers in arXiv include analysis of the probability distributions of words and their combinations, ensuring that they fall into patterns that are consistent with existing subject classes. This serves as a check of the subject categorizations provided by submitters, and helps to detect non-research content.

Fake papers generated by SCIGen software, for example, have a 'native dialect' that can be picked up by simple stylometric analysis (see J. N. G. Binongo *Chance* **16**, 9–17; 2003). The most frequent words used in English text (stop words such as 'the', 'of', 'and') encode stylistic features that are independent of content. On average, these words follow a power-law distribution that is evident in even relatively small amounts of text; significant deviations signal outliers.

The effect can be seen in principal-component analysis plots (see 'Counterfeit clusters'). Computer-generated articles form tight clusters that are well separated from human-authored articles.

Paul Ginsparg *Cornell University, Ithaca, New York, USA.*
ginsparg@cornell.edu

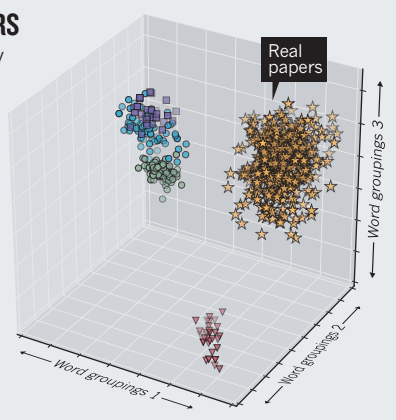
Free up systems for funding and advice

As president of the New Zealand Association of Scientists, I endorse Peter Gluckman's principles for effective science advice to government (*Nature* **507**, 163–165; 2014). As he remarks,

COUNTERFEIT CLUSTERS

Nonsense papers generated by software such as SCIGen and Mathgen cluster separately from human-authored arXiv papers when analysed for stylistic word features.

- SCIGen
- ▼ Mathgen
- SCIGen-physics
- Ike Antkare (SCIGen)
- ★ arXiv 14 March 2014



however, science advisers may encounter a conflict of interest if they are involved in administering public research funding.

Gluckman is the New Zealand Prime Minister's chief science adviser and chaired the panel that last year selected the National Science Challenges. He has been instrumental in publicizing and defending the new funding mechanism for meeting these goals (see go.nature.com/cmgekx1), which the government has signalled are likely to set the default funding strategy for New Zealand science in the next decade and beyond (see, for example, go.nature.com/srrtym).

The community of scientists is concerned about the perceived conflict of interest and loss of trust inherent in combining these roles. They are worried that the challenges will shut out excellent science that does not fit with the goals. Another issue is the perception among Maori researchers that the processes for identifying the national challenges have so far marginalized Maori participation.

It is to be hoped that Gluckman's ten principles will help in future to separate science advisory and funding systems, and that the promised National Statement of Science Investment will address the wider (and no less important) research agenda.

Nicola Gaston *New Zealand Association of Scientists, Wellington, New Zealand.*
president@scientists.org.nz

Journals must boost data sharing

The journal ecosystem is a powerful filter of scientific literature, promoting the best work into the best journals. Why not use a similar mechanism to encourage more comprehensive data sharing?

Several journals have introduced policies mandating that data be shared on a public archive at publication (see, for example, go.nature.com/b7u4ed). However, these policies have met with limited success, perhaps because of authors' fears of losing control, being scooped in subsequent papers or having errors exposed. Moreover, compliance with data-sharing policies is typically checked only after the paper has been accepted.

To spur excellence in data sharing, journals must recognize that better sharing leads to stronger papers, and judge submissions accordingly. Articles associated with feeble sharing efforts should either improve or be rejected.

A focus on publishing verifiable research will boost journal reputation. It also signals to the community of authors that withholding data will restrict them to publication in less-prestigious journals.

Timothy H. Vines *University of British Columbia, Vancouver, Canada.*
vines@zoology.ubc.ca

Projects powered by free computing grid

Herman Tse describes the scientific output of IBM's World Community Grid as "lacklustre" (*Nature* **507**, 431; 2014). This is not the case: the 22 projects we have supported so far have generated more than 35 peer-reviewed papers in prominent journals. Our donated computing power has resulted in several important practical scientific advances.

For example, Japan's Chiba Cancer Center used our free computing power to screen three million drug candidates for treating neuroblastoma, a common childhood cancer. This yielded seven promising compounds that have no apparent side effects (Y. Nakamura *et al. Cancer Med.* **3**, 25–35; 2014).

Last June, Harvard University's Clean Energy Project announced some 35,000 organic materials that could double the efficiency of carbon-based solar cells, after using our grid to scan more than two million candidate materials (see J. Hachmann *et al. Energy Environ. Sci.* **7**, 698–704; 2014, and go.nature.com/cxt181).

Neither should Tse underestimate papers that focus "solely on the technical aspect of distributed computing". Such computing accelerates research and underpins scientific advances. Take the 2013 Nobel Prize in Chemistry: it was awarded to three scientists who developed the kind of computer-modelling techniques on which the work of World Community Grid researchers is based. As the Nobel committee noted: "Today the computer is just as important a tool for chemists as the test tube." **Juan Hindo** *World Community Grid, Chicago, Illinois, USA.*
juan.hindo@us.ibm.com

CONTRIBUTIONS

For Correspondence author guidelines, see go.nature.com/cmchno.

SOURCE: C. LABBÉ

Interneuron subtypes and orientation tuning

ARISING FROM B. V. Atallah, W. Bruns, M. Carandini & M. Scanziani *Neuron* **73**, 159–170 (2012); N. R. Wilson, C. A. Runyan, F. L. Wang & M. Sur *Nature* **488**, 343–348 (2012); S.-H. Lee *et al. Nature* **488**, 379–383 (2012)

Parvalbumin-positive (PV^+) and somatostatin-positive (SST^+) interneurons are two principal subtypes of cortical GABAergic neurons that differ in morphology, physiological properties and postsynaptic targeting^{1–4}. Although GABAergic inhibition is known to be crucial for shaping orientation tuning in the visual cortex^{5–7}, it is unclear whether PV^+ and SST^+ interneurons have different roles. Recently, Atallah *et al.*⁸, Wilson *et al.*⁹ and Lee *et al.*¹⁰ addressed this issue by optogenetically manipulating each interneuron subtype, reaching different conclusions: some investigators found that SST^+ neuron activation sharpens orientation tuning⁹, whereas PV^+ neuron activation has little effect^{8,9}, but others found that activation of PV^+ but not SST^+ interneurons sharpens orientation tuning¹⁰. To understand the cause for the discrepancy we examined the impact of several experimental differences among the studies—anaesthesia and the level and duration of optogenetic stimulation—and found that the discrepancies can be largely explained by differences in the level and duration of interneuron activation. There are replies to this Brief Communication Arising by Atallah, B. V., Scanziani, M. & Carandini, M. *Nature* **508**, <http://dx.doi.org/10.1038/nature13129> (2014) and El-Boustani, S., Wilson, N. R., Runyan, C. A. & Sur, M. *Nature* **508**, <http://dx.doi.org/10.1038/nature13130> (2014).

Atallah *et al.*⁸ showed that mild PV^+ activation caused no significant reduction in orientation tuning width (σ), whereas Lee *et al.*¹⁰ showed that σ reduction is strongly correlated with firing rate reduction (Supplementary Fig. 3e in Lee *et al.*¹⁰). This indicates that the difference between the findings is related to the degree of PV^+ activation. To test the dependence of σ reduction on the level of interneuron activation further, we measured tuning curves of each neuron in anaesthetized mouse V1 at several light intensities (Fig. 1). For PV^+ activation, both the firing rate and σ decreased progressively with increasing light intensity (Fig. 1a, b). Increasing levels of SST^+ activation, however, caused a progressive reduction of firing rate but not of σ (Fig. 1c, d). The range of PV^+ activation-induced firing rate reduction reported by Atallah *et al.*⁸ is indicated by grey shading in Fig. 1b. Within this range we also observed only weak sharpening of orientation tuning, indicating that different levels of PV^+ activation can largely explain the different observations in Lee *et al.*¹⁰ and Atallah *et al.*⁸. Mechanistically, subtractive inhibition by PV^+ activation (Fig. 3 in ref. 10) is considered to be the main cause for the sharpening: stronger PV^+ activation causes a greater increase of spike threshold and therefore stronger firing rate reduction and sharpening.

However, the discrepancy between the findings of Lee *et al.*¹⁰ and Wilson *et al.*⁹ is more profound, and there are two major experimental differences: choice of anaesthetics and laser stimulus duration. As shown in Fig. 2a–e, the specific effect of PV^+ activation on σ reduction was also found in awake, head-fixed mice, indicating that it is not restricted to particular anaesthesia. Meanwhile, in each trial of visual stimulation (4 s), Wilson *et al.*⁹ applied 1 s of laser stimulation at the beginning of visual stimuli, whereas Lee *et al.*¹⁰ applied laser stimulation throughout the 4 s. To test the effect of laser duration, we measured tuning with 1-s laser stimulation in awake PV-channelrhodopsin-2 (ChR2) and SST-ChR2 mice. In both cases, neuronal firing was reduced only during the 1 s of laser stimulation (Fig. 2f, h). In PV-ChR2 mice, 1-s stimulation caused no significant sharpening of tuning, whereas in SST-ChR2 mice it caused a significant sharpening (Fig. 2f–j).

The lack of sharpening with 1-s PV^+ activation is not unexpected given that inhibition was enhanced only during 1 out of 4 s of visual stimulation (Fig. 2f). However, the sharpening by 1 s but not 4 s of

SST^+ activation was unexpected. Notably, during the last 3 s of visual stimulation we observed significant σ reduction with 1-s ($P = 0.003$) but not 4-s ($P = 0.47$) SST^+ activation. One possibility is that whereas the 1-s SST^+ activation caused inhibition of PV^+ neurons^{11,12}, the laser offset caused rebound activity of PV^+ neurons, which in turn caused σ reduction after laser offset.

Finally, we performed cell-attached recordings from genetically labelled cell types without ChR2 stimulation. Pyramidal, PV^+ and SST^+ neurons all exhibited sustained responses to visual stimulation with similar time courses (Fig. 2k), suggesting that both PV^+ and SST^+ neurons are naturally co-active with pyramidal neurons during visual stimulation. Future studies are necessary to understand fully how different spatio-temporal patterns of PV^+ and SST^+ neuron activation affect visual cortical processing, as simultaneous activation or silencing of all neurons belonging to a particular subtype is still a relatively crude manipulation. Nevertheless, these results indicate that the discrepancy between Lee *et al.*¹⁰ and Wilson *et al.*⁹ can be largely explained by different durations of optogenetic activation.

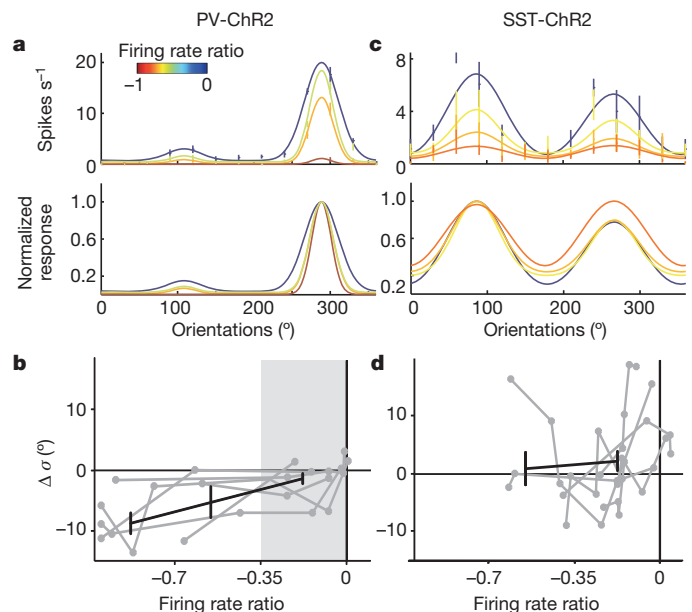


Figure 1 | Correlation between σ reduction and firing rate reduction.

a, b, Juxtacellular recording from putative pyramidal neurons in anaesthetized PV-ChR2 mice showed that σ is reduced progressively with increasing levels of PV^+ neuron activation and pyramidal neuron firing rate reduction. **a**, Top: tuning curves of an example neuron at different levels of optogenetic activation; blue to red, firing rate ratio from 0 to -1 ; error bars indicate s.e.m.; curve, Gaussian fit of the data. Bottom: tuning curves of the same neuron, each normalized by its peak firing rate. **b**, Change of σ versus firing rate ratio. Each grey line represents a single cell; black line, average $\Delta\sigma$ within three ranges of firing rate ratio (0 to -0.35), (-0.35 to -0.7) and (-0.7 to -1). Error bars indicate s.e.m. ($n = 8$ cells); grey shading, range of firing rate ratios reported by Atallah *et al.*⁸. **c, d**, Similar to **a, b**, but for SST-ChR2 mice ($n = 8$).

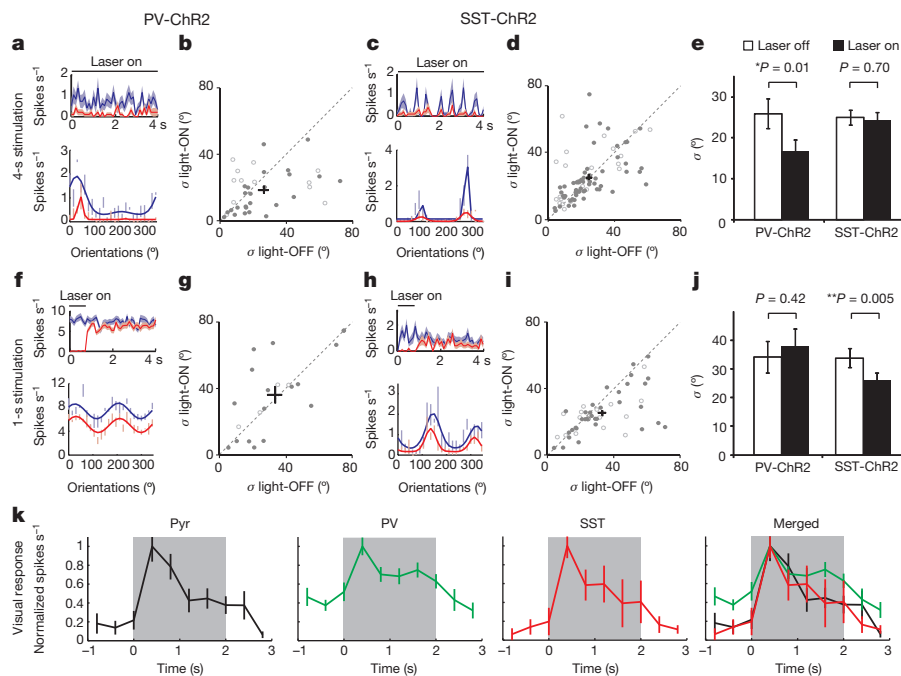


Figure 2 | Effects of 1-s and 4-s laser stimulation on σ in awake PV-ChR2 and SST-ChR2 mice. **a–e**, 4-s activation of PV⁺ but not SST⁺ neurons sharpened orientation tuning. **a**, An example neuron in a PV-ChR2 mouse. Top: firing rates with (red) and without (blue) laser stimulation. Shading indicates s.e.m. Bottom: tuning curves of the same cell with (red) and without (blue) laser stimulation; vertical lines indicate firing rate at each orientation, mean \pm s.e.m. **b**, Population summary of σ in PV-ChR2 mice ($n = 37$ cells). Filled circles, cells with firing rate ratio < 0 ; open circles, cells with firing rate ratio > 0 ; black cross, mean \pm s.e.m. of the population. **c, d**, Similar to **a, b**, but

for SST-ChR2 mice ($n = 81$). **e**, Mean σ across each population of neurons with (black) and without (white) light stimulation (only cells with firing rate ratio < 0 were included). Error bars indicate s.e.m. **f–j**, Similar to **a–e**, but with 1-s laser stimulation at the beginning of each trial. PV-ChR2, $n = 21$; SST-ChR2, $n = 49$. **k**, Time course of visually driven responses of pyramidal ($n = 15$), PV⁺ ($n = 18$) and SST⁺ ($n = 4$) neurons in anaesthetized mice. The response of each neuron was normalized by its peak firing rate and averaged across cells. Grey shading indicates period of visual stimulation. Error bars indicate s.e.m.

Methods

PV-Cre or SST-Cre mice were injected with AAV2/2 for ChR2 expression and implanted with head plates 2–3 weeks before recording. All procedures were approved by Animal Care and Use Committee at UC Berkeley. In anaesthetized mice (PV-ChR2, $n = 8$ mice; SST-ChR2, $n = 8$), juxtacellular recording was made under a two-photon microscope, and optogenetic stimulation was made with LED through the objective. In awake mice (PV-ChR2, $n = 25$; SST-ChR2, $n = 21$), recording was made with multichannel silicon probes. Visual stimulation, optogenetic activation and data analysis followed ref. 10. Tuning width change was measured by $\sigma_{\text{light-ON}} - \sigma_{\text{light-OFF}}$. Firing rate ratio was measured during visual stimulation as $(\text{firing rate}_{\text{light-ON}} - \text{firing rate}_{\text{light-OFF}}) / (\text{firing rate}_{\text{light-ON}} + \text{firing rate}_{\text{light-OFF}})$.

Seung-Hee Lee[†], Alex C. Kwan[†] & Yang Dan¹

¹Division of Neurobiology, Department of Molecular and Cell Biology, Helen Wills Neuroscience Institute, Howard Hughes Medical Institute, University of California, Berkeley, California 94720, USA.

email: ydan@berkeley.edu

[†]Present addresses: Department of Biological Sciences, Korea Advanced Institute of Science and Technology, Daejeon 305-701, Korea (S.-H.L.); Department of Psychiatry, Yale University School of Medicine, New Haven, Connecticut 06511, USA (A.C.K.).

Received 29 August 2013; accepted 28 January 2014.

- Markram, H. *et al.* Interneurons of the neocortical inhibitory system. *Nature Rev. Neurosci.* **5**, 793–807 (2004).
- Ascoli, G. A. *et al.* Petilla terminology: nomenclature of features of GABAergic interneurons of the cerebral cortex. *Nature Rev. Neurosci.* **9**, 557–568 (2008).

- Xu, X., Roby, K. D. & Callaway, E. M. Immunochemical characterization of inhibitory mouse cortical neurons: three chemically distinct classes of inhibitory cells. *J. Comp. Neurol.* **518**, 389–404 (2010).
- Fishell, G. & Rudy, B. Mechanisms of inhibition within the telencephalon: “where the Wild Things are”. *Annu. Rev. Neurosci.* **34**, 535–567 (2011).
- Sillito, A. M. Inhibitory mechanisms influencing complex cell orientation selectivity and their modification at high resting discharge levels. *J. Physiol. (Lond.)* **289**, 33–53 (1979).
- Nelson, S., Toth, L., Sheth, B. & Sur, M. Orientation selectivity of cortical neurons during intracellular blockade of inhibition. *Science* **265**, 774–777 (1994).
- Katzner, S., Busse, L. & Carandini, M. GABA_A inhibition controls response gain in visual cortex. *J. Neurosci.* **31**, 5931–5941 (2011).
- Atallah, B. V., Bruns, W., Carandini, M. & Scanziani, M. Parvalbumin-expressing interneurons linearly transform cortical responses to visual stimuli. *Neuron* **73**, 159–170 (2012).
- Wilson, N. R., Runyan, C. A., Wang, F. L. & Sur, M. Division and subtraction by distinct cortical inhibitory networks *in vivo*. *Nature* **488**, 343–348 (2012).
- Lee, S.-H. *et al.* Activation of specific interneurons improves V1 feature selectivity and visual perception. *Nature* **488**, 379–383 (2012).
- Xu, H., Jeong, H.-Y., Tremblay, R. & Rudy, B. Neocortical somatostatin-expressing GABAergic interneurons disinhibit the thalamorecipient layer 4. *Neuron* **77**, 155–167 (2013).
- Pfeffer, C. K., Xue, M., He, M., Huang, Z. J. & Scanziani, M. Inhibition of inhibition in visual cortex: the logic of connections between molecularly distinct interneurons. *Nature Neurosci.* **16**, 1068–1076 (2013).

Author Contributions S.-H.L. and Y.D. conceived and designed the experiments.

S.-H.L. and A.C.K. performed experiments. S.-H.L., A.C.K. and Y.D. analysed the data. S.-H.L. and Y.D. wrote the manuscript.

Competing Financial Interests Declared none.

doi:10.1038/nature13128

Atallah *et al.* reply

REPLYING TO S.-H. Lee, A. C. Kwan & Y. Dan *Nature* **508**, <http://dx.doi.org/10.1038/nature13128> (2014)

Solving discrepancies in the literature is critical for the advancement of science, and the Comment by Lee *et al.*¹ is thus welcome. It clarifies that there is no contradiction between the earlier study of Lee *et al.*² and our study³. The disagreement is in the interpretation of the results and in the model used to fit the data.

Both studies^{2,3} optogenetically perturbed the activity of inhibitory parvalbumin-expressing (PV⁺) cells in mouse visual cortex and measured the resulting impact on the orientation tuning width of pyramidal cells. Whereas Lee *et al.*² reported a narrowing of the tuning width, our study³ did not observe any systematic change.

The most obvious effect on photoactivation of PV⁺ cells is the reduction in the firing of pyramidal cells. Accordingly, we are glad that when Lee *et al.*¹ (see accompanying Comment) consider the same range of pyramidal cell firing reduction as that described in our study³ (that is, up to ~50% reduction, for a ratio of -0.34), there is no narrowing of the tuning curve. This narrowing, on the other hand, is present when exploring larger reductions in pyramidal cell firing, consistent with the findings of Lee *et al.*².

These effects can be explained by the simple linear model with threshold proposed by our study³. In this model, the impact of PV⁺ cells is to subtract and scale orientation tuning curves, unless the firing rate is 0. The model captures not only the data in our study³, but also the data points of the example cell shown in figure 1 of the Comment¹. As illustrated in our Fig. 1, this model fits the data very well, so all these results can be explained by a simple ‘iceberg effect’.

Moreover, the model explains additional data obtained by our study³, which are not mentioned in the Comment¹. Our study³ performed the reverse experiment, namely the optogenetic suppression of PV⁺ cells to increase pyramidal cell firing rate up to 250%, and again found that there was no systematic increase in tuning width. As described in our study³, the model explains this finding because once the iceberg is out of the water it cannot get wider by rising further.

Therefore, one can fully reconcile Lee *et al.*² and our study³ by pointing out (1) that when one explores both intermediate and large reductions in pyramidal cell firing rates one sees both effects (invariance and narrowing of tuning width, respectively, as the Comment does¹); and (2) that when one uses the linear-threshold model³ one explains all of these effects. We believe that interpreting the data in the context of such a model is superior to comparing Gaussian functions fit to responses obtained with and without stimulation, as was done in Lee *et al.*². It is closer to the biological reality of a spike threshold, more parsimonious,

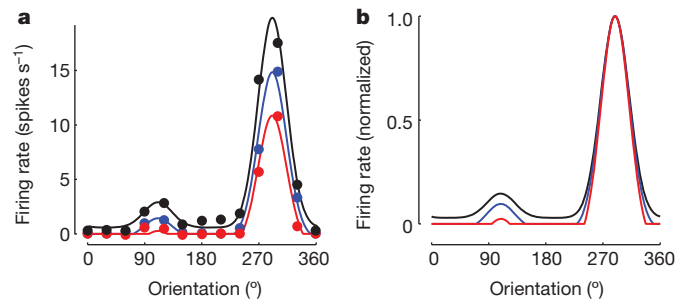


Figure 1 | The linear-threshold model captures the effects of increasing PV activation. **a**, Responses of a pyramidal cell to stimuli of different orientations, in control conditions (black) or in the presence of increasing PV cell activation (blue, moderate PV cell activation; red, stronger PV cell activation). The data are the same of those in figure 1 of Lee *et al.*¹, and were obtained with the Matlab function ‘grabit.m’. We did not consider a fourth set of points, with the lowest firing rate, as there are barely any data points above zero. The curves indicate the fits of the linear-threshold model introduced by our study³. **b**, The same curves as in **a**, rescaled to peak at 1, to illustrate a mild but progressive narrowing of tuning curves with increasing PV⁺ cell activation.

and therefore more informative as to the functional effect of PV⁺ cells on pyramidal cells.

Bassam V. Atallah^{1,2}, Massimo Scanziani² & Matteo Carandini³

¹Champlimaud Neuroscience Programme, Lisbon 1400-038, Portugal.

²Howard Hughes Medical Institute and Center for Neural Circuits and Behavior and Neurobiology Section, Division of Biology, University of California San Diego, La Jolla, California 92093-0634, USA.

email: matteo@cortexlab.net

³UCL Institute of Ophthalmology, University College London, London EC1V 9EL, UK.

1. Lee, S.-H., Kwan, A. C. & Dan, Y. Interneuron subtypes and orientation tuning. *Nature* **508**, <http://dx.doi.org/10.1038/nature13128> (2014).
2. Lee, S.-H. *et al.* Activation of specific interneurons improves V1 feature selectivity and visual perception. *Nature* **488**, 379–383 (2012).
3. Atallah, B. V., Bruns, W., Carandini, M. & Scanziani, M. Parvalbumin-expressing interneurons linearly transform cortical responses to visual stimuli. *Neuron* **73**, 159–170 (2012).

doi:10.1038/nature13129

El-Boustani *et al.* reply

REPLYING TO S.-H. Lee, A. C. Kwan & Y. Dan *Nature* **508**, <http://dx.doi.org/10.1038/nature13128> (2014)

Several recent studies have examined the function of parvalbumin-expressing (PV⁺) and somatostatin-expressing (SST⁺) inhibitory neurons in V1 (refs 1–3). Although it is commonly agreed that these cell types alter the responses of pyramidal neurons in distinct ways—via divisive or subtractive inhibition—their specific roles remain a matter of debate. The Comment by Lee *et al.*⁴ presents new data suggesting that the differences between the results of Lee *et al.*² compared to Atallah *et al.*³ and Wilson *et al.*¹ could be explained by the strength and duration of laser stimulation used to optogenetically activate these

two classes of inhibitory neuron. The data presented by Lee *et al.*⁴ now clarify that PV⁺ neurons, when probed with small amounts of optogenetic activation, do not significantly change the tuning of their target cells, confirming Atallah *et al.*³ and Wilson *et al.*¹. The new SST⁺ results presented in the Comment⁴ show that SST⁺ neurons can subtract responses, consistent with Wilson *et al.*¹, but we suggest that the switch of function of SST⁺ neurons in their data between short (1 s) and long (4–5 s) stimulation reveals a core principle of inhibition in cortical networks rather than simply being a peculiarity of stimulation protocols.

The fundamental difference between these two conditions resides in the temporal overlap between inhibitory neuron activation and target-cell responses: when these overlap, inhibition is divisive (causing no change in tuning width of target neurons), but when they do not overlap, inhibition is subtractive (and reduces tuning width).

Lee *et al.*² found that activation of PV⁺ neurons can reduce the orientation tuning width of target neurons, whereas Atallah *et al.*³ and Wilson *et al.*¹ reported a divisive effect, mainly resulting in scaling down the orientation tuning curve without significant changes of tuning width. If PV⁺ neurons are strongly stimulated with channelrhodopsin-2 (ChR2) by increasing the laser intensity or duration, they can markedly reduce pyramidal neuron responses and narrow their tuning widths, as explicitly noted by Wilson *et al.* (see Supplementary Figs 6 and 7 of ref. 1)—both Wilson *et al.*¹ and Atallah *et al.*³ used only moderate amounts of PV⁺ activation to avoid the ‘floor effect’ on target neurons. The new data presented by Lee *et al.*⁴ (see Fig. 1) now show that moderate stimulation of PV⁺ neurons does not result in substantial tuning width reduction, similar to the findings reported by Atallah *et al.*³ and Wilson *et al.*¹. However, in this stimulation intensity range, they do not report any gain modulation of target cell responses as observed by Wilson *et al.*¹ and Atallah *et al.*³. This discrepancy, as well as the nature of the effect observed in the new data presented by Lee *et al.*⁴, remains to be clarified.

Regarding SST⁺ neurons, both Wilson *et al.*¹ and the new data presented by Lee *et al.*⁴ show that the effect of these neurons on target cells

as measured after 1 s of laser stimulation is subtractive. This is in contrast to the divisive effect found by Lee *et al.*² when SST⁺ neurons were stimulated during the entire duration of visual stimulation. Lee *et al.*⁴ suggest that the new findings may be explained by SST⁺ to PV⁺ inhibition and rebound of PV⁺ neuron activity after laser offset. We believe that this is unlikely, as direct (presumably stronger) activation of PV⁺ inhibition by 1 s of laser stimulation was not strong enough to reduce the tuning width, as reported by Wilson *et al.*¹ and now by Lee *et al.*⁴.

Instead, we suggest that this result can be explained by the response properties of different inhibitory neurons and their co-activation with target cells. The visual responses of SST⁺ neurons are distinct from PV⁺ or pyramidal neurons in terms of response latency⁵, size tuning curve⁶ and correlation with nearby pyramidal cells⁷. Probing the function of inhibitory neurons using visual (as well as ChR2) stimulation invokes these response modes, which shapes their effect on pyramidal cells. Crucially, PV⁺ neurons are routinely co-activated with pyramidal neurons by the visual stimuli used to measure tuning; such co-activation (even when ChR2-mediated inhibition is added in small amounts) fundamentally divides pyramidal neuron responses. The inhibition that SST⁺ neurons provide is a function of whether or not they are co-activated with target pyramidal neurons by visual stimuli and ChR2 (Fig. 1). Prolonged ChR2 stimulation of SST⁺ neurons may force them to be co-active with pyramidal neurons during visual stimulation and the analysis time period (Lee *et al.*²), resulting in a divisive effect. Driving these neurons only during the initial segment of visual response, with pulse trains (Wilson *et al.*¹) or steady stimulation (Lee *et al.*⁴), may reduce the co-activation during the analysis period, resulting in a subtractive effect. More generally, sustained ChR2 activation of inhibitory neuron subtypes is superimposed not only on visual responses of neurons but also on complex inhibitory–disinhibitory circuits⁸, making it difficult to resolve clearly the function of these subtypes *in vivo*. As the duration of ChR2-mediated stimulation of specific inhibitory neurons seems to determine the effect on target cells, the use of moderate single-pulse probes of light could reveal the core principles of direct inhibition in functioning visual cortex networks without the risk of fundamentally perturbing the ongoing network dynamics through recurrent inhibitory circuits⁹.

Sami El-Boustani¹, Nathan R. Wilson¹, Caroline A. Runyan¹ & Mriganka Sur¹

¹Department of Brain and Cognitive Sciences, Picower Institute for Learning and Memory, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA.
email: msur@mit.edu

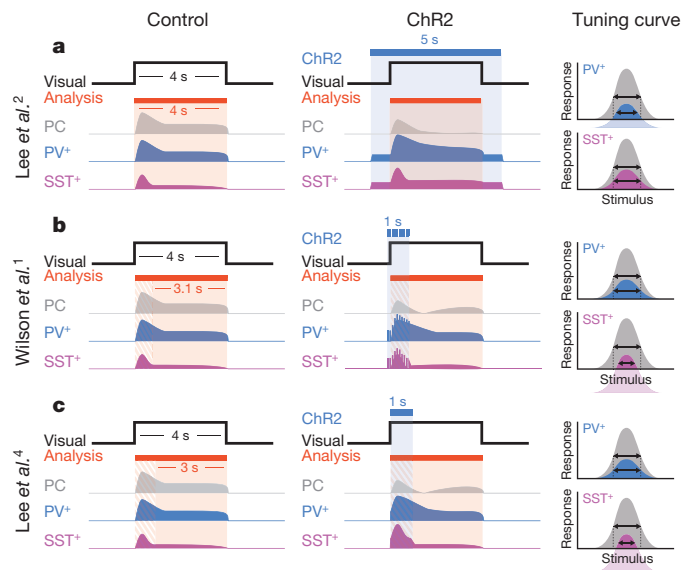


Figure 1 | Schematic of the different experimental protocols used in Lee *et al.*², Wilson *et al.*¹ and Lee *et al.*⁴. a–c, For each study (a, Lee *et al.*²; b, Wilson *et al.*¹; c, Lee *et al.*⁴ (accompanying Comment)), the left panel describes the control protocol where neuron activity is recorded during visual stimulation alone from pyramidal cells (PC), parvalbumin-expressing cells (PV⁺) and somatostatin-expressing cells (SST⁺); the middle panel describes the same experiment when inhibitory neurons are activated with ChR2; and the right panel shows the resulting effect of ChR2 activation on target cell tuning curves for PV⁺ (blue, top) and SST⁺ (pink, bottom) neurons. In the protocol description (left and middle panels), the visual stimulus duration (4 s) is indicated in black, the time epoch during which the analysis has been performed (4, 3 or 3.1 s) in red and the duration of ChR2 activation in blue (middle panel). The hatched red area indicates segment of the visual stimulation that was not always included in the analysis. Note that pulses are indicated by single blue lines (b) whereas continuous stimulation is indicated by a bar (a, c). Stereotypical responses of pyramidal neurons (grey) as well as PV⁺ (blue) and SST⁺ (pink) inhibitory neurons are depicted (left and middle panels), as well as the suggested effect of temporal co-activation on pyramidal neuron tuning width (arrows, right panel).

- Wilson, N. R., Runyan, C. A., Wang, F. L. & Sur, M. Division and subtraction by distinct cortical inhibitory networks *in vivo*. *Nature* **488**, 343–348 (2012).
- Lee, S.-H. *et al.* Activation of specific interneurons improves V1 feature selectivity and visual perception. *Nature* **488**, 379–383 (2012).
- Atallah, B. V., Bruns, W., Carandini, M. & Scanziani, M. Parvalbumin-expressing interneurons linearly transform cortical responses to visual stimuli. *Neuron* **73**, 159–170 (2012).
- Lee, S.-H., Kwan, A. C. & Dan, Y. Interneuron subtypes and orientation tuning. *Nature* **508**, <http://dx.doi.org/10.1038/nature13128> (2014).
- Ma, W. *et al.* Visual representations by cortical somatostatin inhibitory neurons—selective but with weak and delayed responses. *J. Neurosci.* **30**, 14371–14379 (2010).
- Adesnik, H., Bruns, W., Taniguchi, H., Huang, Z. J. & Scanziani, M. A neural circuit for spatial summation in visual cortex. *Nature* **490**, 226–231 (2012).
- Gentet, L. J. *et al.* Unique functional properties of somatostatin-expressing GABAergic neurons in mouse barrel cortex. *Nature Neurosci.* **15**, 607–612 (2012).
- Pfeffer, C. K., Xue, M., He, M., Huang, Z. J. & Scanziani, M. Inhibition of inhibition in visual cortex: the logic of connections between molecularly distinct interneurons. *Nature Neurosci.* **16**, 1068–1076 (2013).
- Brunel, N. & Hakim, V. Fast global oscillations in networks of integrate-and-fire neurons with low firing rates. *Neural Comput.* **11**, 1621–1671 (1999).

Author Contributions S.E.-B. did additional experiments, and S.E.-B. and M.S. wrote the response with contributions from N.R.W. and C.A.R.

doi:10.1038/nature13130

CANCER

Disabling defences in the brain

The identification of the enzyme plasmin as a defence against cancer cells that have spread to the brain, and of tumour-cell serpin proteins that inhibit plasmin production, outlines a mechanism for the formation of brain metastases.

JANINE T. ERLER

Most deaths from cancer are not caused by the primary tumour. A cancer can spread from the primary tumour to other organs through a process known as metastasis, and it is the growth of metastatic tumours that ultimately compromises normal organ function and is responsible for more than 90% of deaths in cancer patients¹. Brain metastases present one of the poorest prognoses for cancer patients, and their rate of incidence is increasing². Fortunately, metastasis is a highly inefficient process: fewer than 0.01% of cells that leave a primary tumour are able to colonize and grow in other organs³. The underlying molecular mechanisms that govern initial metastatic cancer-cell survival and growth in secondary organs remain largely unknown and are an area of intensive research. Writing in *Cell*, Valiente *et al.*⁴ show how cancer cells that have metastasized to the brain overcome death signals from host tissue cells and use the pre-existing vasculature to enable their proliferative growth.

To form brain metastases, cancer cells must not only cross the blood–brain barrier — a cellular interface that selectively controls which molecules can pass between the bloodstream and the brain — but also survive in this new, very different environment, and obtain sufficient blood supply for proliferation into tumours. It has been known for some time that most cancer cells that manage to cross the blood–brain barrier die in the brain tissue⁵, and that those cells that survive are tightly associated with pre-existing capillary blood vessels⁶, using them as a blood supply in a process known as vascular co-option. Valiente *et al.* are the first to identify the mechanisms underlying these crucial steps.

Using cultures of mouse brain slices and *in vivo* mouse models of brain metastasis, the authors show that proteins called anti-plasminogen activator (PA) serpins play a crucial part in allowing cancer cells to overcome death signals released by astrocyte cells in the brain. Anti-PA serpins are a family of protease-enzyme inhibitors that prevent the conversion of the molecule plasminogen into the enzyme plasmin⁷. The authors found that plasmin expressed by astrocytes not only induces

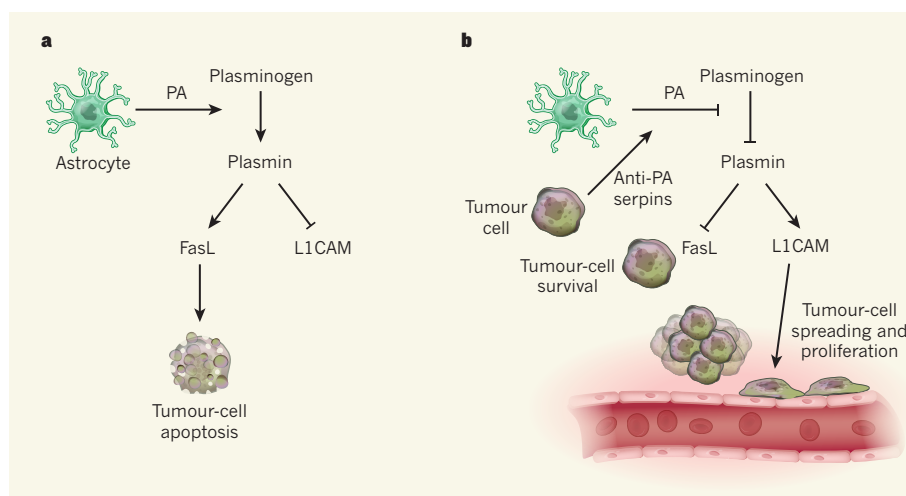


Figure 1 | Plasmin protection. **a**, Astrocytes in the brain produce plasminogen activator (PA) proteins that induce the production of the enzyme plasmin from its precursor plasminogen. Valiente *et al.*⁴ show that plasmin both induces the death of cancer cells that have spread to the brain, through action of the protein Fas ligand (FasL), and inhibits cancer-cell spread along capillaries, by inactivating the protein L1CAM. **b**, However, they also find that some cancer cells that have reached the brain express high levels of anti-PA serpin proteins, and suggest that the activity of these proteins aids the survival and spreading of metastatic cancer cells along blood vessels, enabling their proliferative growth and the formation of lethal metastatic tumours.

apoptotic cell death (mediated by the protein Fas ligand) in cancer cells, but also prevents the cancer cells from binding to and spreading along brain capillaries, by inactivating the axon-pathfinding molecule L1CAM (Fig. 1).

Valiente and colleagues' analysis of gene expression in brain-seeking lung and breast cancer cells revealed that various anti-PA serpins (most commonly neuroserpin and serpin B2) are highly expressed in such cells compared with non-metastatic cells from the primary tumour and also with metastatic cells targeting other organs. The researchers demonstrate that the action of these anti-PA serpins in preventing the function of plasmin allows cancer-cell survival and vascular co-option. In support of these findings, they show that expression of neuroserpin and serpin B2 in primary human lung tumours was associated with metastasis to the brain but not to the bone or other lung sites, and that these anti-PA serpins were highly expressed in patients' brain metastases.

This elegant study highlights the complex interplay between cancer cells and host (stromal) cells that is needed for metastasis.

Moreover, it demonstrates the intrinsic ability of cancer cells to adapt to inhospitable micro-environments. The work not only provides the first molecular insight into how cancer cells successfully colonize and grow in the brain, but also sheds light on how vascular co-option takes place, a process that has remained obscure despite being known as a key step in the metastasis of many cancer types.

The study has several limitations. For example, the authors use an intracardiac mouse model of organ colonization, in which cells cultured *in vitro* are slowly injected into the bloodstream through the heart. Although this model is the best way of driving tumour-cell colonization of the brain, it does not recapitulate the whole metastatic cascade. Thus, 'brain-tropic' tumour cells may use subtly different and, more importantly, multiple molecular mechanisms for brain colonization, survival and metastatic growth. Another issue is that the authors predominantly use human cancer-cell lines in mice that lack intact immune systems, and therefore do not fully recapitulate normal tumour–stromal cell interactions. However, they were limited by the fact that suitable

models of spontaneous brain metastasis do not exist, and they did confirm several of their findings in mice with intact immune systems, using mouse cancer-cell lines.

Metastasis from lung and breast cancers, as studied in this work, accounts for around two-thirds of human brain-metastasis cases, but it remains to be tested whether the mechanisms that the authors have uncovered also underlie brain metastases from other cancer types, such as melanoma. It will also be interesting to assess how these mechanisms relate to primary brain cancers, in particular because lowering of L1CAM expression levels has been shown to induce apoptosis in cells of glioma tumours⁸. And, building a broader view of the molecular network outlined by the authors — for example, by analysing the role of integrin proteins, which bind to L1CAM⁹ and regulate cell spreading, or by assessing other genes upregulated in brain-tropic cancer cells — may provide a deeper understanding of the regulation of brain metastasis¹⁰.

A further aspect to consider is whether new therapies could result from this study. Valiente *et al.* show that blocking the function of Fas ligand can protect cancer cells from death in the brain. Thus, it would be interesting to test whether the addition of soluble Fas ligand might sensitize metastatic cells to apoptotic death and whether this pathway could be used as a therapy. Moreover, might targeting certain PA proteins, such as tissue PA or urokinase PA, be an effective approach by which to combat metastasis, especially given that these are already targets of interest for anticancer therapy?

Anti-PA serpins, plasmin, Fas ligand and L1CAM are all associated with poor patient prognosis in many types of cancer, suggesting that Valiente and colleagues' findings not only have advanced our understanding of metastasis to the brain, but also may have uncovered a more general mechanism of metastatic spread. It will be exciting to see whether the molecular players they identify

are important in other types of cancer, and whether similar tumour–stromal cell interactions permit metastatic growth in organs other than the brain. ■

Janine T. Erler is at the Biotech Research and Innovation Centre (BRIC), University of Copenhagen, Copenhagen 2200, Denmark. e-mail: janine.erler@bric.ku.dk

1. Sporn, M. B. *Lancet* **347**, 1377–1381 (1996).
2. Maher, E. A., Mietz, J., Arteaga, C. L., DePinho, R. A. & Mohla, S. *Cancer Res.* **69**, 6015–6020 (2009).
3. Mehlen, P. & Puisieux, A. *Nature Rev. Cancer* **6**, 449–458 (2006).
4. Valiente, M. *et al. Cell* **156**, 1002–1016 (2014).
5. Heyn, C. *et al. Magn. Reson. Med.* **56**, 1001–1010 (2006).
6. Kienast, Y. *et al. Nature Med.* **16**, 116–122 (2010).
7. Croucher, D. R., Saunders, D. N., Lobov, S. & Ranson, M. *Nature Rev. Cancer* **8**, 535–545 (2008).
8. Bao, S. *et al. Cancer Res.* **68**, 6043–6048 (2008).
9. Felding-Habermann, B. *et al. J. Cell Biol.* **139**, 1567–1581 (1997).
10. Erler, J. T. & Lindig, R. J. *Pathol.* **220**, 290–296 (2010).

IMMUNOLOGY

A is for immunity

If a pregnant mouse lacks vitamin A, her offspring are born with smaller lymph nodes and have impaired immune responses as adults — a finding that adds immune development to this vitamin's list of key functions. SEE LETTER P.123

GÉRARD EBERL

Vitamin A has fundamental roles in embryonic development, reproduction, bone metabolism, skin homeostasis and vision. Therefore, vitamin A deficiency has profound effects on the fetus and in early childhood, and is responsible for the high prevalence of childhood blindness and for many deaths in regions plagued with malnutrition. Another major role for vitamin A is in immunity — this has been known about for more than 40 years, but only fully investigated in the past decade. On page 123 of this issue, van de Pavert *et al.*¹ report that vitamin A provided by a mother mouse to her fetus is required for the offspring to develop full-size lymph nodes, the organs in which adaptive immune responses are generated. As a consequence, newborn mice that have not received optimal vitamin A exposure *in utero* are more susceptible to infection.

In the early 1970s, it was discovered² that chickens deprived of vitamin A have low numbers of lymphocytes (cells of the adaptive immune system, including T and B cells) and that injection of vitamin A boosts specific antibody responses and skin-graft rejection in mice³, indicating that the vitamin influences immune activity. Later work⁴

showed that vitamin A regulates the balance of different types of lymphocyte-effector cell: supplementation of vitamin A in mice favours the emergence of type 2 T-helper cells, which

drive immunity to parasitic worms, over type 1 T-helper cells, which drive antiviral responses.

Research into the impact of vitamin A on immunology really took off when it was found that retinoic acid, a product of vitamin A metabolism, regulates the migration and differentiation of T cells in the intestine. Dendritic cells, which activate T cells and determine their activation and migration patterns, were found⁵ to express the enzyme retinaldehyde dehydrogenase, which converts retinol (the form of vitamin A required for vision) into retinoic acid. Binding of this metabolite to the retinoic-acid receptor (RAR), expressed by T cells, results in enhanced expression of

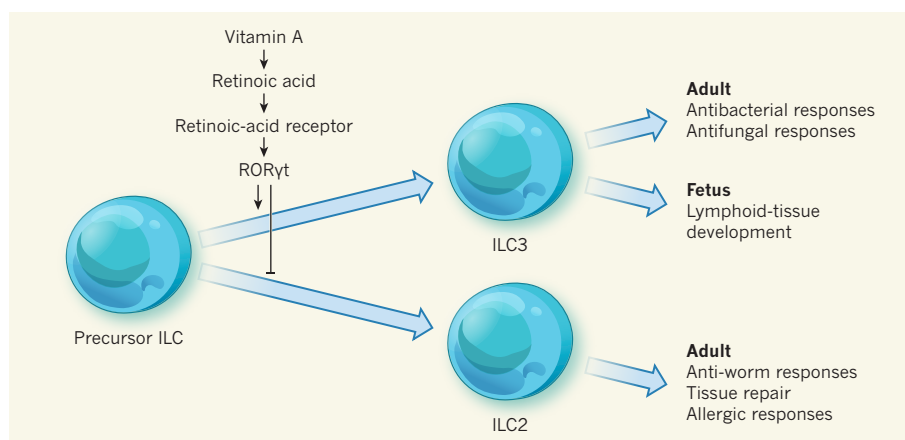


Figure 1 | Vitamin A regulates innate and adaptive immune responses. Van de Pavert *et al.*¹ show that the vitamin A metabolite retinoic acid and its receptor contribute to the expression of the hormone receptor RORγt, which is needed¹¹ for the generation and maturation of type 3 innate lymphoid cells (ILC3s) from their precursor cells. In adults, ILC3s initiate inflammatory responses against microbial pathogens; in the fetus, they induce the development of lymphoid tissues by mimicking these inflammatory responses. Through its influence on this pathway, vitamin A deficiency has several effects on immunity: in adults, it leads to the preferential generation of type 2 innate lymphoid cells (ILC2s), which are involved in anti-worm immunity, tissue repair and allergic responses⁹; in the fetus, van de Pavert *et al.* show that lack of the vitamin causes the development of undersized lymphoid tissues and deficient adaptive immune responses after birth.

surface molecules that induce the T cells to migrate to the intestinal lamina propria⁵ — a layer of the body's mucosal linings. Retinoic acid also promotes the generation of regulatory T cells⁶, which control inflammatory immune responses, and of pro-inflammatory T cells in a diseased intestinal environment⁷.

Earlier this year, another effect of vitamin A on the immune system was uncovered. Fighting a virus, a bacterium, a worm or a wound requires different immune responses that are orchestrated by distinct types of T cell. It takes several days for the right T cells to be selected, instructed and multiplied in the lymph nodes, and then dispatched to the endangered tissue. But another cell family, termed innate lymphoid cells (ILCs), can do a similar job within hours, responding directly to cues from signalling molecules that are produced in the infected or injured tissues⁸. Just a few months ago, it was shown⁹ that vitamin A deficiency causes this response to shift from ILCs associated with the pro-inflammatory antibacterial activity of type 3 immunity (ILC3s), to an anti-worm and tissue-repair (type 2) response (Fig. 1). This finding suggested that malnutrition redirects the immune system to protect against large parasites that might compete for food. And it hinted at an even more fundamental role for vitamin A in influencing immunity, because ILC3s are also the cells that induce the development of lymph nodes.

One of the authors of the present study, with colleagues, had previously discovered that an unusual cell type accumulates in the developing lymph nodes of the fetus¹⁰. Generation of these cells, called lymphoid tissue inducer (LTi) cells, requires the hormone receptor ROR γ t (ref. 11), which is also needed for the generation of type 3 T cells¹² and ILC3s⁸. It has since become clear that LTi cells are prototypical ILC3s that function early in antimicrobial immune responses, and that their activity is harnessed in the fetus to induce lymph-node development through a process that essentially mimics an inflammatory response against pathogens.

Now, van de Pavert and colleagues' findings suggest that fetal LTi cells cannot be generated without retinoic acid. The authors show that retinoic acid induces the maturation of LTi cells from their precursor cells in developing lymph nodes by inducing RAR to bind the promoter region of the gene *Rorc*, which encodes ROR γ t, thereby initiating its expression. In mice that lacked functional RAR, mature LTi cells were less numerous and expressed lower levels of pro-inflammatory molecules, the lymph nodes were significantly smaller and the adaptive immune response to viral infection was blunted, compared with mice that had normal RAR activity.

Dendritic cells, demonstrated to be the source of retinoic acid in immune cells in adults, cannot be the source in the fetus because dendritic cells colonize developing lymph nodes only around the time of birth.

Some of the authors of the current study had previously suggested that retinoic acid contributes to lymph-node development¹³, and proposed that retinoic acid in the fetus is synthesized by nerves, which express high levels of retinaldehyde dehydrogenase. Now, van de Pavert *et al.* show that lymph-node size and the efficiency of immune responses are impaired in mice whose mothers were fed vitamin-A-deficient diets during pregnancy — suggesting the possibility that the mother is responsible for both the absorption of vitamin A from food and its conversion to retinoic acid.

In light of these results, the lasting effects of malnutrition on public health cannot be overestimated. Maternal and childhood vitamin A deficiency may not only cause severe physical handicaps but also undermine the efficacy of childhood vaccination campaigns in already impoverished regions. Pregnant women who smoke or drink are at heightened risk of passing on immune-development deficiencies to their children, because both smoking and alcohol consumption are associated with vitamin A deficiency. Conversely, diets too rich in vitamin A may contribute to the increase

in inflammatory conditions that is occurring in the industrialized world. Thus, it seems that a careful balance of vitamin A — dietary or supplemented — is desirable. ■

G rard Eberl is in the Lymphoid Tissue Development Unit, Institut Pasteur, 75724 Paris, France.
e-mail: gerard.eberl@pasteur.fr

1. van de Pavert, S. A. *et al.* *Nature* **508**, 123–127 (2014).
2. Bang, B. G., Bang, F. B. & Foard, M. A. *Am. J. Pathol.* **68**, 147–162 (1972).
3. Jurin, M. & Tannock, I. F. *Immunology* **23**, 283–287 (1972).
4. Cantorna, M. T., Nashold, F. E. & Hayes, C. E. *J. Immunol.* **152**, 1515–1522 (1994).
5. Iwata, M. *et al.* *Immunity* **21**, 527–538 (2004).
6. Mucida, D. *et al.* *Science* **317**, 256–260 (2007).
7. DePaolo, R. W. *et al.* *Nature* **471**, 220–224 (2011).
8. Spits, H. & Di Santo, J. P. *Nature Immunol.* **12**, 21–27 (2011).
9. Spencer, S. P. *et al.* *Science* **343**, 432–437 (2014).
10. Mebius, R. E., Rennert, P. & Weissman, I. L. *Immunity* **7**, 493–504 (1997).
11. Eberl, G. *et al.* *Nature Immunol.* **5**, 64–73 (2004).
12. Ivanov, I. I. *et al.* *Cell* **126**, 1121–1133 (2006).
13. van de Pavert, S. A. *et al.* *Nature Immunol.* **10**, 1193–1199 (2009).

This article was published online on 19 March 2014.

SOLAR SYSTEM

Ring in the new

Planets are no longer the only Solar System bodies sporting ring systems. Two dense rings have been detected encircling a Centaur object — a relatively small, icy interloper from the distant reaches of the Solar System. [SEE LETTER P.72](#)

JOSEPH A. BURNS

To date, rings containing icy and rocky particles have been known to orbit only the four giant planets in the Solar System, most notably Saturn¹. Now, continuing nature's excellent track record of upending the firm beliefs of planetary theorists, a fully developed ring system has been detected surrounding the Centaur object (10199) Chariklo, as Braga-Ribas *et al.*² report on page 72 of this issue.

Centaur objects are relatively small icy bodies that travel along short-lived, moderately elliptical orbits in the region of the giant planets. The discovery was made when Braga-Ribas and colleagues stationed an armada of small telescopes across a 1,500-kilometre-wide swathe of South America to watch as the enigmatic Chariklo crossed in front of a star (Fig. 1). Astronomers use such stellar occultations just as a motorist gauges an obstacle's location and silhouette on a dimly lit highway by watching background lights flicker as the obstacle passes before them. Here, the authors sought to use this technique to refine Chariklo's size and to determine its shape during a predicted

occultation. But they — and we — got much more.

Chariklo's rings are not the first to be discovered by occultation. Although planetary experts³ had maintained for decades that special circumstances had led Saturn, alone among the planets, to sport rings, a complex ring system was detected^{4,5} nestled within a few radii of Uranus during an occultation in 1977. In the following decade, many occultations probed Neptune's surroundings, but only occasionally was the star observed to flicker, and then on just one side of the planet⁶. Eventually, the realization struck that the hit-and-miss nature of these few successful detections could be explained if Neptune's rings were restricted to localized arcs covering about 10% of the rings' circumference⁷. Images obtained with the Voyager 2 spacecraft confirmed⁸ this unexpected morphology in 1989.

This is not the first time that Chariklo has surprised observers. Following its discovery in 1997, the object's brightness systematically and mysteriously dropped by 40% (ref. 9), and its strong water-ice signature gradually faded¹⁰, only to have these trends reverse since 2008. The identification of Chariklo's rings now

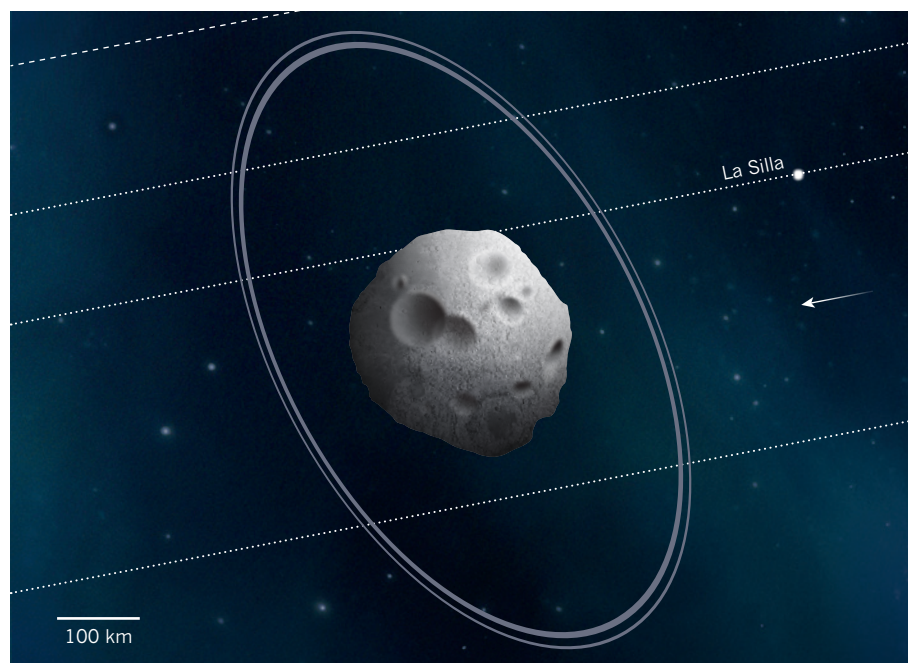


Figure 1 | Chariklo's rings. When a Solar System body transits in front of a star, blocking its light, the star will blink off for an interval proportional to the length of the chord across the body. Observing such a stellar occultation, Braga-Ribas *et al.*² detected two dark rings surrounding Chariklo, a Centaur object with a radius of about 125 kilometres. The inner ring, centred at 391 km, is about 7 km across and extinguishes roughly 40% of the star's light. A clear gap of about 9 km separates it from the outer ring, which is centred at 405 km, is roughly 3 km wide and removes about 5% of the starlight. The star's track (dotted lines) crossed previously unknown rings as seen from several separate sites, including the La Silla Observatory in Chile, which obtained the highest time resolution. The star's path (dashed line) seen from another observatory missed the rings. Chariklo is darker than drawn here. (Illustration adapted from Fig. 2 of the paper².)

explains these puzzling decreases in brightness and spectral intensity: they occur as ice-rich rings, which have 15% of Chariklo's surface area but 3 times its reflectivity, become edge-on when viewed from Earth. Like so many baffling observations in science, the answer is obvious once its explanation is known.

Of the four planetary ring systems, that of Uranus provides the closest analogue to Chariklo's pair. Its dozen or so isolated, coal-black rings — most only a few kilometres wide, with crisp edges and sometimes slightly variable widths — are separated by clear gaps. Small moons known as shepherd satellites were suggested¹¹ to prise open lanes in the Uranian rings and to confine their edges. In this mechanism, high-order gravitational perturbations of moons generate repulsive torques on nearby disks of orbiting material.

To be effective at their task, the shepherds need to have masses comparable to those of the ringlets they herd. In other words, the moons should be perhaps a few kilometres in radius — too small to be visible through a ground-based telescope. When Voyagers 1 and 2 subsequently spied two tiny moons herding Saturn's F ring, another pair guiding particles in Uranus' ϵ ring, and also the small moon Pan in the Encke gap in Saturn's A ring, this clever mechanism became enshrined as fact¹. It is now routinely called upon to explain openings and confined rings in all sorts of astrophysical

disks, even in the protoplanetary disks within which planets form.

Naturally, then, Braga-Ribas and co-authors hypothesize that hidden small shepherd satellites account for the gap between Chariklo's rings as well as the crisp peripheries of both rings. But a dirty secret of planetary rings should be exposed: following exhaustive searches since 2004 using the Cassini spacecraft, it is almost certain that none of the numerous gaps in Saturn's C ring and in its Cassini Division (a low-density band between Saturn's main A and B rings) harbour any shepherds of the requisite size¹².

Perhaps the physics missing in our attempts to explain such ring features as gaps will be revealed by investigations of Chariklo's much simpler system. Its rings extend a mere three-thousandth the dimensions of Saturn's; indeed, its entire retinue could slip — with much room to spare — through the largest gaps at the outer and inner edges of Saturn's rings. Circular orbital speeds in the ring regions surrounding the planets and Chariklo scale with the size of the central body, and are merely tens of metres per second for Chariklo's ring particles. Relative speeds between adjacent particles in Chariklo's rings will allow collisions that thicken the rings, whereas Saturn's rings, with their higher orbital speeds, are much thinner. The reduced gravity, low speeds and resulting gentle collisions at Chariklo permit ring

dynamics to be investigated in a previously unimagined regime.

How might such a diminutive ring system have formed? Given Chariklo's relatively small gravitational influence, it seems improbable that its rings formed contemporaneously with the processes that gave birth to the Centaur itself. A more likely scenario is one similar to that proposed for the origin of our Moon¹³, in which a nearly catastrophic collision between the primordial Earth and a Mars-sized object lofted copious impact ejecta into an orbiting disk, from which the Moon subsequently formed. Whatever its cause, as this disk spreads out by a few Chariklo radii, the largest shards in the disk could shepherd the remaining disk material.

An alternative model arises from the fact that about five per cent of Centaurs and Trans-Neptunian Objects have small companions, possibly obtained as a result of three-body interactions¹⁴. Such satellites may be gently disrupted by impacts of interplanetary debris. At Chariklo's ring distances, orbital speeds are low, but some of the impact ejecta will depart from their source even more slowly, implying that a ring-like tube of fragments might enshroud the satellite's orbit¹⁵.

The detection of rings around Chariklo will startle many planetary theorists. But so it has always been in planetary exploration: theoretical ideas rarely generate searches that lead to discovery — rather, discoveries such as this prompt us to new understandings. ■

Joseph A. Burns is in the Departments of Astronomy and of Mechanical and Aerospace Engineering, Cornell University, Ithaca, New York 14853, USA.
e-mail: joseph.burns@cornell.edu

1. Tiscareno, M. S. in *Planets, Stars and Stellar Systems* Vol. 3 (eds Oswalt, T., French, L. M. & Kalas, P.) 309–376 (Springer, 2013).
2. Braga-Ribas, F. *et al. Nature* **508**, 72–75 (2014).
3. Pollack, J. B. *Space Sci. Rev.* **18**, 3–93 (1975).
4. Elliot, J. L., Dunham, E. & Mink, D. *Nature* **267**, 328–330 (1977).
5. Millis, R. L., Wasserman, L. H. & Birch, P. V. *Nature* **267**, 330–331 (1977).
6. Roques, F. *et al. Bull. Am. Astron. Soc.* **16**, 1027–1028 (1984).
7. Hubbard, W. B. *et al. Nature* **319**, 636–640 (1986).
8. Smith, B. A. *et al. Science* **246**, 1422–1449 (1989).
9. Bel'skaya, I. N. *Icarus* **210**, 472–479 (2010).
10. Guilbert, A. *et al. Astron. Astrophys.* **501**, 777–784 (2009).
11. Goldreich, P. & Tremaine, S. *Nature* **277**, 97–99 (1979).
12. Colwell, J. E. *et al. in Saturn From Cassini-Huygens* (eds Dougherty, M. K., Esposito, L. W. & Krimigis, S. M.) 375–412 (Springer, 2009).
13. Canup, R. M. & Righter, K. (eds) *Origin of the Earth and Moon* (Univ. Arizona Press, 2000).
14. Noll, K. S., Grundy, W. M., Chiang, E. J., Margot, J.-L. & Kern, S. D. in *The Solar System beyond Neptune* (eds Barucci, M. A., Boehnhardt, H., Cruikshank, D. P., Morbidelli, A. & Dotson, R.) 345–363 (Univ. Arizona Press, 2008).
15. Burns, J. A. *et al. Science* **284**, 1146–1150 (1999).

This article was published online on 26 March 2014.

Brain vessels squeezed to death

A study finds that contractile cells that surround the capillary vessels of the brain control the blood supply to healthy neurons, and that their death may aggravate brain injury by strangling vessels. [SEE ARTICLE P.55](#)

DANIEL M. GREIF & ANNE EICHMANN

The human brain's two billion neurons are supplied with oxygen and nutrients by a network of cerebral blood vessels that equates to around 700 kilometres in length. Most of these blood vessels are tiny capillaries surrounded by contractile supporting cells called pericytes. When a major vessel that supplies blood to the brain becomes obstructed, capillary blood flow is interrupted, and this may deplete the brain's oxygen supply, causing rapid loss of neuronal function, cell death and stroke. In this issue, Hall *et al.*¹ (page 55) show that pericytes have a crucial role in the normal regulation of cerebral blood flow, and that a lack of oxygen causes pericyte death and constriction of capillaries.

Around 140 years ago, the French physiologist Charles Rouget observed² a population of cells lying in close proximity to capillary cells, which he hypothesized were involved in vessel contraction. Although initially termed Rouget's cells, they were subsequently named pericytes, to describe their location around capillaries. Of all the organs in the body, the

brain has the highest density of pericytes. Brain pericytes are known to be needed for the development and maintenance of the blood–brain barrier^{3–5}, which regulates exchanges between the brain and its vasculature, and to be involved in diseases associated with this interface⁶. However, the additional role for pericytes first suggested by Rouget — regulating cerebral blood flow — has not been demonstrated in the brain *in vivo* until now.

Neuronal activity stimulates blood supply by causing vessel dilation, a fact that is exploited by functional magnetic resonance imaging to allow mapping of the brain regions that are associated with different tasks. To determine the role of pericytes in vasodilation, Hall *et al.* first investigated the molecular signals regulating capillary dilation induced by neuronal activity in slices of rodent cerebellum. The authors found that pericytes relax their grip on capillaries in response to glutamate molecules, the key excitatory neurotransmitters in the brain. This causes the capillaries to dilate, directing more blood to active brain regions (Fig. 1a).

Next, the researchers investigated which branches of the cerebral vasculature are

responsible for dynamic alterations in vessel diameter and blood flow. For example, is dilation of larger arterioles, induced by neuronal activity, sufficient to cause an increase in blood flow in smaller capillaries⁷, or is vasodilation initiated at capillaries? Previous research⁸ from the same group demonstrated that electrical stimulation of pericytes (which mimics neural activity) in isolated rodent retinas alters capillary diameter at the pericytes.

In the present study, Hall and colleagues used *in vivo* live imaging of vessels in the somatosensory cortex region of the mouse brain — which responds to sensory inputs — after electrical stimulation of the animal's whisker pad. In doing so, they demonstrated that relaxation of pericytes at the capillary level precedes relaxation of smooth-muscle cells (which surround larger arterioles and are involved in cerebral blood-vessel constriction) by about 1 second. Thus, pericytes surrounding capillaries initiate changes in local blood flow in response to neuronal activity, and thereby provide a sensitive control mechanism to match neuronal activity and blood supply throughout the brain. In support of the idea that capillaries are a primary site of cerebral blood-flow control, the terminals of one neuronal subtype (noradrenergic neurons) are typically found in proximity to capillaries, rather than near to arterioles⁹.

Hall *et al.* report that the resting diameter of capillaries is larger in areas where pericytes are present than in areas that lack these cells. Electrically stimulated dilations were more frequent and of larger magnitude in capillary areas with pericytes. Previously, embryonic mice that lack brain pericytes have been shown to develop microaneurysms¹⁰. Together, these results indicate that pericytes play a key part in the regulation of blood-vessel size. It would be informative to remove or reduce pericyte

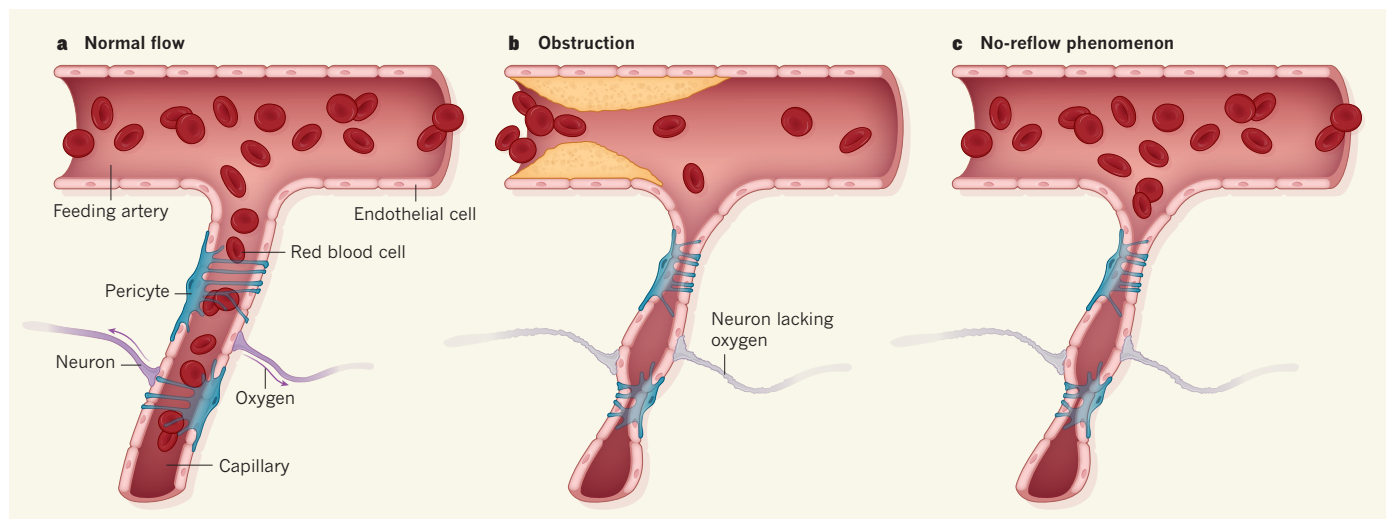


Figure 1 | Cutting off the blood supply to the brain. Hall *et al.*¹ report that pericytes, which surround the endothelial cells that form capillaries, are involved in regulating cerebral blood flow. **a**, Blood flows into capillaries from feeding arteries. In response to neuronal activity, pericytes relax their grip on cerebral capillaries, causing blood-vessel dilation and increasing the oxygen supply to active neurons. **b**, Obstruction of a cerebral artery (which mimics the effects of ischaemic stroke) results in pericyte constriction and death, limiting blood flow and oxygen supply to neurons in the brain. **c**, The authors hypothesize that the no-reflow phenomenon occurs because of rigor mortis in dead pericytes, which causes continued strangulation of capillaries even after a blocked artery is reopened.

numbers in the central nervous system of adult mice, or to block the pericyte's ability to contract. Subsequent analysis of the changes in cerebral capillary blood flow induced by neuronal activity in these animals compared with controls would help to delineate a more precise role for pericytes in regulating blood flow in the brain.

Ischaemic stroke, caused by a lack of blood to the brain, has devastating effects. Treatments are limited because reperfusion strategies that aim to restore blood supply are effective within only around 4.5 hours of the onset of symptoms. Treatment efficacy is limited by time because reperfusion is impaired when the reopening of a large artery is delayed (known as the no-reflow phenomenon). Although controversial, persistent pericyte contraction after ischaemia has been implicated as a cause of the no-reflow phenomenon in capillaries^{11,12}. Hall and colleagues demonstrate that, on exposure of rat brain slices to conditions simulating ischaemia, capillaries constrict and then pericytes die (Fig. 1b). Furthermore, the authors found that in rats, temporary obstruction of a cerebral artery *in vivo* induces substantial death of pericytes, but not of endothelial

cells that make up the blood vessels.

Hall and co-workers suggest that the long-term reduction of cerebral blood flow after reperfusion of a blocked artery may be at least partly attributable to pericyte rigor mortis — literally 'the stiffness of death' (Fig. 1c). In rigor mortis, stiffness results from a lack of ATP molecules, which prevents myosin and actin — two proteins that interact to cause muscle contraction — from being separated from one another.

The concept of prolonged vasoconstriction due to strangulation by dead pericytes raises several questions that require further investigation. Why are pericytes, as opposed to other cell types such as endothelial cells, particularly susceptible to ischaemia-induced death? Can pericyte death be prevented, and will this inhibit the no-reflow phenomenon? Is pericyte rigor mortis a factor in this phenomenon after injury in other tissues, particularly in the heart following a heart attack? If this model of ischaemia-induced pericyte rigor mortis remains robust in the face of further analysis, it may pave the way for approaches to combat ischaemic injury that prevent pericyte-induced loss of a tissue's ability to reperfuse. ■

Daniel M. Greif and Anne Eichmann are at the Yale Cardiovascular Research Center, Yale University School of Medicine, New Haven, Connecticut 06510, USA. A.E. is also in the Department of Cellular and Molecular Physiology, Yale University School of Medicine, and at the Center for Interdisciplinary Research in Biology, Collège de France, Paris. e-mails: daniel.greif@yale.edu; anne.eichmann@yale.edu

1. Hall, C. N. *et al.* *Nature* **508**, 55–60 (2014).
2. Rouget, C. *Arch. Physiol. Normal Pathol.* **5**, 603–663 (1873).
3. Daneman, R., Zhou, L., Kebede, A. A. & Barres, B. A. *Nature* **468**, 562–566 (2010).
4. Armulik, A. *et al.* *Nature* **468**, 557–561 (2010).
5. Bell, R. D. *et al.* *Neuron* **68**, 409–427 (2010).
6. Bell, R. D. *et al.* *Nature* **485**, 512–516 (2012).
7. Fernández-Klett, F., Offenhauser, N., Dirnagl, U., Priller, J. & Lindauer, U. *Proc. Natl Acad. Sci. USA* **107**, 22290–22295 (2010).
8. Peppiatt, C. M., Howarth, C., Mobbs, P. & Attwell, D. *Nature* **443**, 700–704 (2006).
9. Cohen, Z., Molinatti, G. & Hamel, E. *J. Cereb. Blood Flow Metab.* **17**, 894–904 (1997).
10. Lindahl, P., Johansson, B. R., Levéen, P. & Betsholtz, C. *Science* **277**, 242–245 (1997).
11. Yemisci, M. *et al.* *Nature Med.* **15**, 1031–1037 (2009).
12. Vates, G. E., Takano, T., Zlokovic, B. & Nedergaard, M. *Nature Med.* **16**, 959 (2010).

This article was published online on 26 March 2014.

PLANETARY SCIENCE

A chronometer for Earth's age

Simulations of Earth's growth show a correlation between the timing of the Moon's formation and the amount of mass that Earth accreted afterwards. This relationship provides a way of measuring the age of our planet. [SEE LETTER P.84](#)

JOHN CHAMBERS

The age of the oldest objects in the Solar System is known with remarkable precision — 4,567 million years^{1,2} — thanks to recent strides in the dating of meteorites. Unfortunately, applying the same dating methods to Earth yields an age that is frustratingly fuzzy. Our planet formed sometime during the first 150 million years of the Solar System's history^{3–5}, but we do not know when. On page 84 of this issue, Jacobson *et al.*⁶ propose a way of measuring the time at which Earth finished forming using numerical simulations of the planet's growth and its chemical composition. Their result: Earth formed in 95 million years, with an uncertainty of about 32–39 million years, making the planet about 4,470 million years old.

A big part of the problem with dating Earth is that our planet did not appear overnight. Starting from humble beginnings, Earth gradually accumulated material over an extended

period of time. Indeed, it is still gaining mass today in the form of meteorites and interplanetary dust particles. What is needed is a milestone in Earth's growth at which we can say that the planet was essentially complete. A widely adopted milestone is the major collision with a planet-sized body that is thought to have formed the Moon⁷ (Fig. 1). Current theory suggests that Earth experienced several of these 'giant impacts' during its formation, with the Moon-forming impact being the last. Each impact mixed together material that would end up in Earth's metallic core and rocky mantle, as well as adding bulk to both.

The prolonged nature of Earth's growth and the existence of multiple giant impacts complicate the task of dating Earth using radiometric clocks — those that combine known decay rates of radioactive materials with measurements of how these materials and their decay products are distributed within Earth today. For one thing, it is unclear just how much mixing of core and mantle material occurred

with each giant impact, and to what extent the radiometric clocks were reset as a result.

Earth probably retains a memory of several of these events, rather than just the last one. Giant impacts may have ejected some of the planet's more volatile elements into space, distorting radiometric dating systems that assume that all of the radioactive decay products are still present. It is also possible that Earth's core and mantle continued to interact for some time after the last giant impact.

This is where modelling the growth of our planet could pay dividends. The formation of the Sun's rocky planets probably passed through several stages⁸, beginning with micrometre-sized dust grains in the nascent Solar System, proceeding to asteroid-sized bodies known as planetesimals, and then to a few dozen Moon-to-Mars-mass planetary embryos. The accumulation of these planetary embryos into the modern rocky planets through occasional giant impacts was by far the slowest stage, and it largely determined how long Earth took to form.

Jacobson *et al.* have modelled this final growth stage, beginning with various populations of planetary embryos and planetesimals. They also examined two widely different scenarios for what the giant planets of the Solar System were doing during this time. En route, the authors found an interesting and remarkably robust correlation: the timing of the last giant impact on Earth is inversely related to the amount of mass the planet accumulated afterwards from leftover planetesimals. If the last giant impact occurred early on, there would have been plenty of planetesimals left



Figure 1 | Earthrise from the Moon. Jacobson *et al.*⁶ find that the timing of the Moon-forming impact on Earth is inversely related to the amount of mass that the planet accumulated afterwards.

for Earth to sweep up afterwards. If the last giant impact was late, few planetesimals would have remained, and Earth's growth would have largely ceased.

This correlation provides an independent way of dating the Moon-forming impact, provided that we can measure the amount of material that arrived subsequently. Fortunately, there is a way to do this. Several elements, such as iridium and platinum, show a strong tendency to move into Earth's core. During the upheaval of each giant impact, these elements leached from the planet's mantle, bonding with heavy, iron-rich material destined to sink to the core. After the last giant impact, Earth's mantle should have been almost completely stripped of iridium, platinum and their cousins.

In practice, these elements are present in small amounts in the mantle, and in the same relative proportions seen in many meteorites⁹. To many researchers, this suggests that Earth acquired a fraction of its mass after the last giant impact, when the core and mantle had ceased separating. Jacobson *et al.* combine the measured mass of this extra material with the correlation from their simulations to date the Moon-forming impact and then deduce Earth's age. They find it very unlikely that Earth finished forming in the first 38 million years of the Solar System. Their favoured time — 95 million years — is compatible with some radiometric-dating estimates for when Earth's core finished forming⁴, and makes Earth comfortably older than the oldest minerals known to have formed in its crust¹⁰.

Naturally, Jacobson and colleagues' method is only as valid as our picture of how planets form. The standard model for planet formation in the Solar System is far from complete, and it has had a torrid time lately trying to explain some aspects of extrasolar planetary systems, such as the existence of Earth-sized

planets orbiting very close to their star¹¹. If conditions during the growth of the planets were different from those assumed by Jacobson *et al.*, then the authors' estimated age for Earth could be incorrect.

Despite this caveat, it is encouraging to see studies that combine the fundamental physics inherent in numerical simulations of planet formation with the wealth of information available on Earth's composition. Understanding how and when the Sun's planets formed is immensely challenging, and researchers need every tool available. Studies such as the one by Jacobson and colleagues may be the best hope for understanding how and when our planet came to be. ■

John Chambers is in the Department of Terrestrial Magnetism, Carnegie Institution for Science, Washington DC 20015, USA. e-mail: jchambers@carnegiescience.edu

1. Amelin, Y. *et al. Earth Planet. Sci. Lett.* **300**, 343–350 (2010).
2. Connelly, J. N. *et al. Science* **338**, 651–655 (2012).
3. Yin, Q. *et al. Nature* **418**, 949–952 (2002).
4. Allège, C. J., Manhès, G. & Göpel, C. *Earth Planet. Sci. Lett.* **267**, 386–398 (2008).
5. Halliday, A. N. *Phil. Trans. R. Soc. A* **366**, 4163–4181 (2008).
6. Jacobson, S. A. *et al. Nature* **508**, 84–87 (2014).
7. Canup, R. M. & Asphaug, E. *Nature* **412**, 708–712 (2001).
8. Morbidelli, A., Lunine, J. I., O'Brien, D. P., Raymond, S. N. & Walsh, K. J. *Annu. Rev. Earth Planet. Sci.* **40**, 251–275 (2012).
9. Walker, R. J. *Chem. Erde Geochem.* **69**, 101–125 (2009).
10. Valley, J. W. *et al. Nature Geosci.* **7**, 219–223 (2014).
11. Dong, S. & Zhu, Z. *Astrophys. J.* **778**, 53 (2013).

CANCER

Clonal cooperation

Widespread genetic heterogeneity in cells of human tumours poses a question: what prevents the fittest clone from taking over? A demonstration of interdependence between distinct clones might shed light on this puzzle. [SEE LETTER P.113](#)

KORNELIA POLYAK & ANDRIY MARUSYK

Sequencing of human cancer genomes has revealed a high degree of genetic heterogeneity among cells of a given tumour¹. In most cases, tumour growth is thought to be driven by the most 'advanced' cancer-cell subpopulation — that carrying the highest number of cancer-driving mutations. However, the presence of many mutations that occur at only low frequency implies that tumours contain multiple subclones, and the relevance of these is not fully understood. On page 113 of this issue, Cleary *et al.*² provide a potential explanation for some forms of intratumoral heterogeneity, by describing a

cooperative cellular interaction in mouse mammary tumours in which the presence of two types of clone is required for tumour formation.

Mammary tumours induced by overexpression of the *Wnt1* gene are thought to originate in mammary epithelial stem cells³, which can differentiate into both the luminal and basal cells that make up mammary epithelial tissue. These tumour-initiating cells can therefore give rise to tumours composed of cancer cells with either basal or luminal features. It was previously established⁴ that the coexistence of these two lineages is maintained by paracrine interactions between the two cell types — that is, short-distance interactions mediated by signalling molecules — because only luminal

cells produce Wnt1, and basal cells depend on this signalling protein to proliferate. The first report⁵ of *Wnt1*-induced mammary tumours noted that many tumours contain more than one clone, but the mechanistic basis of this heterogeneity had not been explored.

Now, Cleary *et al.* find that, in a subset of *Wnt1*-induced mammary tumours, the basal cells harbour a potent cancer-driving mutation in the *Hras* gene that is not detected in the cancerous luminal cells. This suggests that the two differentiation lineages in the tumours represent different tumour clones. Both of these lineage-restricted clones are essential for tumour formation, because neither was able to generate tumours by itself when transplanted into secondary recipient mice, whereas the combined transplantation of both cell types was highly tumorigenic (Fig. 1).

To study this interclonal cooperation further, the authors used transgenic mice in which *Wnt1* expression is regulated by a gene promoter that requires the small molecule doxycycline for activity; in this model, the tumours regress following doxycycline withdrawal. Transplantation of cells from these inducible tumours into the mammary fat pads of wild-type mice or mice in which *Wnt1* is continuously expressed led to tumour generation in the presence of doxycycline in both cases. However, the transplanted tumours behaved differently in the two recipient strains after doxycycline withdrawal. Complete regression occurred in the wild-type animals, although small numbers of residual tumour cells survived and led to tumour recurrence after doxycycline readministration. By contrast, only partial regression was observed in mice continuously expressing *Wnt1*, and the

initial regression was followed by rapid tumour regrowth, despite the continued absence of doxycycline.

Most surprisingly, the authors found a large number of *Wnt1*-transgene-expressing luminal cells from the recipient mice in these recurrent tumours, suggesting that the void left by loss of *Wnt* expression in tumour luminal cells can be filled by recruiting *Wnt*-expressing wild-type luminal cells. As with the parental tumours, both *Hras*-mutant basal cells and *Hras*-wild-type luminal cells were required for efficient tumour maintenance.

Although this study provides convincing evidence for cooperation between basal and luminal clones and for the requirement of both cell types in *Wnt1*-driven tumours, extrapolating these findings to explain subclonal intratumoral heterogeneity in human breast cancer is not straightforward. The reasons for this are that, first, the authors used a transgenic mouse model in which the transgene is expressed in a large number of mammary epithelial cells. Thus, biclonal tumours could have evolved following tumour initiation in two independent cells that happened to be next to each other, rather than reflecting two subclones derived from a single initiating cell. The likelihood of two independently initiated transformed cells forming a single tumour mass in humans is very low, but the overall conclusions of the study regarding clonal cooperation are still valid.

Second, the paracrine interdependence between basal and luminal cells, which enables independent evolution within the two differentiation lineages in this mouse cancer model, might represent a special case that is not necessarily relevant to primary human

cancers. Although the *Wnt1* pathway is one of the main cancer-causing pathways, *Wnt1* activation has rarely been observed in human breast cancers, and most such tumours do not depend on *Wnt1* activity for growth. Paracrine signalling between basal and luminal cells plays a key part in normal mammary-gland development^{6,7}, but the degree to which these interactions are preserved in tumours is uncertain, especially given that most human breast tumours are thought to originate in luminal progenitor cells. Furthermore, because human breast tumours can take decades to evolve, the extent of intratumoral heterogeneity in such tumours is much higher^{8,9} than that in animal models. Although the existence of paracrine interdependence relieves competition between the lineage-restricted subclones, lack or loss of strict interdependence is expected to allow the fittest clone to take over the tumour. Thus, other mechanisms besides paracrine interdependence between clones may be responsible for intratumoral heterogeneity in human breast cancers.

Despite these differences between the model and human cancers, Cleary and colleagues' study will fuel the already-intense exploration of the functional and clinical relevance of intratumoral heterogeneity. More importantly, their results raise questions about the validity of assays for identifying tumour-initiating cells (also known as cancer stem cells) in human tumours that rely on transplanting single cells or homogeneous cell populations¹⁰. Such assays would inevitably fail to determine the tumorigenic potential of clones that require cooperation with other cells. Many different experimental approaches, in combination with detailed molecular characterization of large cohorts of human tumours at different stages, will be required to answer the many lingering questions regarding intratumoral heterogeneity. Let us hope that acquiring a better understanding of this phenomenon will not only be of scientific interest, but also lead to improved treatment for cancer patients. ■

Kornelia Polyak and Andriy Marusyk are in the Department of Medical Oncology, Dana-Farber Cancer Institute, and the Department of Medicine, Harvard Medical School, Boston, Massachusetts 02115, USA. e-mail: kornelia_polyak@dfci.harvard.edu

- Greenman, C. *et al.* *Nature* **446**, 153–158 (2007).
- Cleary, A. S., Leonard, T. L., Gestl, S. A. & Gunther, E. J. *Nature* **508**, 113–117 (2014).
- Li, Y. *et al.* *Proc. Natl Acad. Sci. USA* **100**, 15853–15858 (2003).
- Kim, S., Goel, S. & Alexander, C. M. *PLoS ONE* **6**, e19310 (2011).
- Mester, J., Wagenaar, E., Sluysers, M. & Nusse, R. *J. Virol.* **61**, 1073–1078 (1987).
- Rosen, J. M. & Roarty, K. *Breast Cancer Res.* **16**, 202 (2014).
- Forster, N. *et al.* *Dev. Cell* **28**, 147–160 (2014).
- Almendo, V. *et al.* *Cell Rep.* **6**, 514–527 (2014).
- Almendo, V. *et al.* *Cancer Res.* **74**, 1338–1348 (2014).
- Meacham, C. E. & Morrison, S. J. *Nature* **501**, 328–337 (2013).

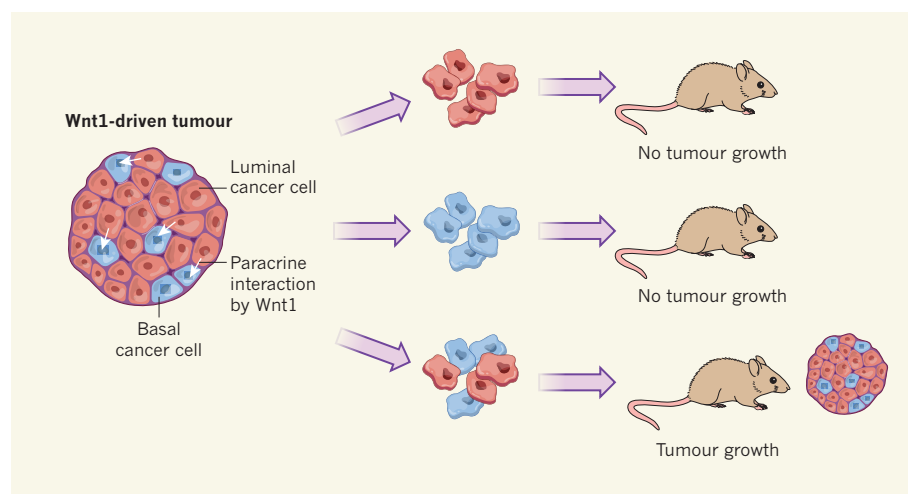


Figure 1 | Luminal and basal clones cooperate to form mammary tumours. Mammary tumours that develop in mice expressing the *Wnt1* transgene are composed of heterogeneous mixtures of basal and luminal cancer cells. Cleary *et al.*² show that, in some such tumours, neither luminal nor basal cancer cells alone can form a new tumour when injected into the mammary fat pads of tumour-prone mice, whereas mixtures of the two cell populations induce biclonal tumour formation with high efficiency. In such tumours, the basal cells also carry cancer-driving mutations in the *Hras* gene. Thus, it seems that paracrine interactions between the *Hras*-mutant basal cells and the luminal cells, which provide the *Wnt1* on which the basal cells depend, is required for the development of these cancers.

Capillary pericytes regulate cerebral blood flow in health and disease

Catherine N. Hall^{1*}, Clare Reynell^{1*}, Bodil Gesslein^{2*}, Nicola B. Hamilton^{1*}, Anusha Mishra^{1*}, Brad A. Sutherland³, Fergus M. O'Farrell¹, Alastair M. Buchan³, Martin Lauritzen^{2,4} & David Attwell¹

Increases in brain blood flow, evoked by neuronal activity, power neural computation and form the basis of BOLD (blood-oxygen-level-dependent) functional imaging. Whether blood flow is controlled solely by arteriole smooth muscle, or also by capillary pericytes, is controversial. We demonstrate that neuronal activity and the neurotransmitter glutamate evoke the release of messengers that dilate capillaries by actively relaxing pericytes. Dilation is mediated by prostaglandin E₂, but requires nitric oxide release to suppress vasoconstricting 20-HETE synthesis. *In vivo*, when sensory input increases blood flow, capillaries dilate before arterioles and are estimated to produce 84% of the blood flow increase. In pathology, ischaemia evokes capillary constriction by pericytes. We show that this is followed by pericyte death in rigor, which may irreversibly constrict capillaries and damage the blood-brain barrier. Thus, pericytes are major regulators of cerebral blood flow and initiators of functional imaging signals. Prevention of pericyte constriction and death may reduce the long-lasting blood flow decrease that damages neurons after stroke.

Pericytes are isolated contractile cells on capillaries that may regulate cerebral blood flow^{1,2}, in addition to stabilizing newly formed capillaries³, maintaining the blood-brain barrier^{4–6}, contributing to the 'glial scar' in pathology⁷, and having stem cell properties⁸. Pericytes can be constricted and dilated by neurotransmitters *in vitro*^{1,9}, through poorly understood signalling pathways, and capillary blood flow heterogeneity might reflect differences in pericyte tone^{10–12}. Pericytes can constrict *in vivo*, but it was suggested² that they do not relax actively to increase blood flow¹³. Similar controversy surrounds the effect of pericytes on blood flow in pathology^{14,15}. We have now characterized the responses of pericytes in the neocortex and cerebellum to neuronal activity and ischaemia. We demonstrate that, unexpectedly, pericytes are the first vascular elements to dilate during neuronal activity, making them the initiators of functional imaging signals. Furthermore, they die readily in ischaemia, and this is expected to promote brain damage.

Signals that regulate pericyte dilation

We assessed the signalling systems that dilate molecular layer capillaries in cerebellar slices¹, using 95% or 20% O₂ in the superfusate to produce a supra-normal or a physiological O₂ concentration, [O₂], in the slice^{16,17}. Capillaries were defined as vessels of less than 10 µm diameter and lacking a continuous layer of smooth muscle. Their diameter was larger ($P = 0.00023$) at the lower [O₂] (5.36 ± 0.30 µm (mean \pm s.e.m., $n = 59$) in 20%, and 4.07 ± 0.14 µm ($n = 154$) in 95% O₂). Capillary pericytes can be identified by labelling for the chondroitin sulphate proteoglycan NG2 or the PDGFR β receptor (Fig. 1a), or using mice expressing DsRed under control of the NG2 promoter¹⁸ (Fig. 1b, c), and are a different cell class from Iba1-expressing perivascular microglia or macrophages¹⁹ (Fig. 1c). For neocortical capillaries in postnatal day 21 (P21) rats the pericyte soma density was 2.2 ± 0.2 per 100 µm of capillary length (950 µm of capillary analysed in each of 11 confocal stacks). Pericytes extend processes along and around vessels (Fig. 1a–c), which regulate capillary diameter.

The capillary diameter was extremely stable over the course of 60 min if no drugs were applied (see below). Application of noradrenaline (2 µM), to mimic its release from the locus coeruleus *in vivo*, produced a sustained constriction mediated by pericytes¹ (Fig. 1d, e), that was not affected by [O₂] (Fig. 1f). Superimposing glutamate (500 µM), to mimic neuronal glutamate release¹³, dilated capillaries at pericyte locations¹ (Fig. 1d, g and Supplementary Video 1). As a percentage of the diameter without drugs, this dilation was twice as large with 20% as with 95% O₂ (Fig. 1h), possibly due to less production of vasoconstricting 20-HETE (20-hydroxyeicosatetraenoic acid) in low [O₂] (see below and Extended Data Fig. 1). Independent of [O₂], most pericytes (72% in 95% O₂; 70% in 20% O₂) constricted >5% to noradrenaline, and dilated >5% to glutamate (63% in 95% O₂ ($n = 154$); 66% in 20% O₂ ($n = 59$)), and most pericytes that constricted to noradrenaline also dilated to glutamate (71% in 95% O₂; 78% in 20% O₂). Glutamate also dilated some capillaries in the absence of noradrenaline: in 20% O₂, 29% (7/24) of pericytes dilated >5%, which is less ($P = 0.005$, chi-squared test) than the 66% of pericytes that dilated >5% after applying noradrenaline. In subsequent experiments noradrenaline was used to pre-constrict capillaries, to aid analysis of the signalling underlying glutamate-evoked dilation.

These experiments do not establish which cells the noradrenaline and glutamate act on (they may be neurons, astrocytes or pericytes themselves¹³) to release the downstream messengers that control pericyte tone. However, we can rule out the idea that noradrenaline generates 20-HETE to constrict pericytes, because blocking 20-HETE synthesis with 1 µM HET0016 did not affect the noradrenaline-evoked constriction (Extended Data Fig. 2b, g).

Glutamate releases nitric oxide (NO), a vasodilator, by activating NMDA (*N*-methyl-D-aspartate) receptors¹³ and an NO donor (DETA-NONOate, 100 µM) evoked capillary dilation (Extended Data Fig. 2m). Blocking NO synthase with L-N^G-nitroarginine (L-NNA, 100 µM) reduced the glutamate-evoked dilation (Fig. 1i, l; $P = 0.002$ from analysis of

¹Department of Neuroscience, Physiology and Pharmacology, University College London, Gower Street, London, WC1E 6BT, UK. ²Department of Neuroscience and Pharmacology and Center for Healthy Aging, University of Copenhagen, DK-2200 Copenhagen N, Denmark. ³Acute Stroke Programme, Radcliffe Department of Medicine, University of Oxford, Oxford OX3 9DU, UK. ⁴Department of Clinical Neurophysiology, Glostrup University Hospital, DK-2600 Glostrup, Denmark.

*These authors contributed equally to this work.

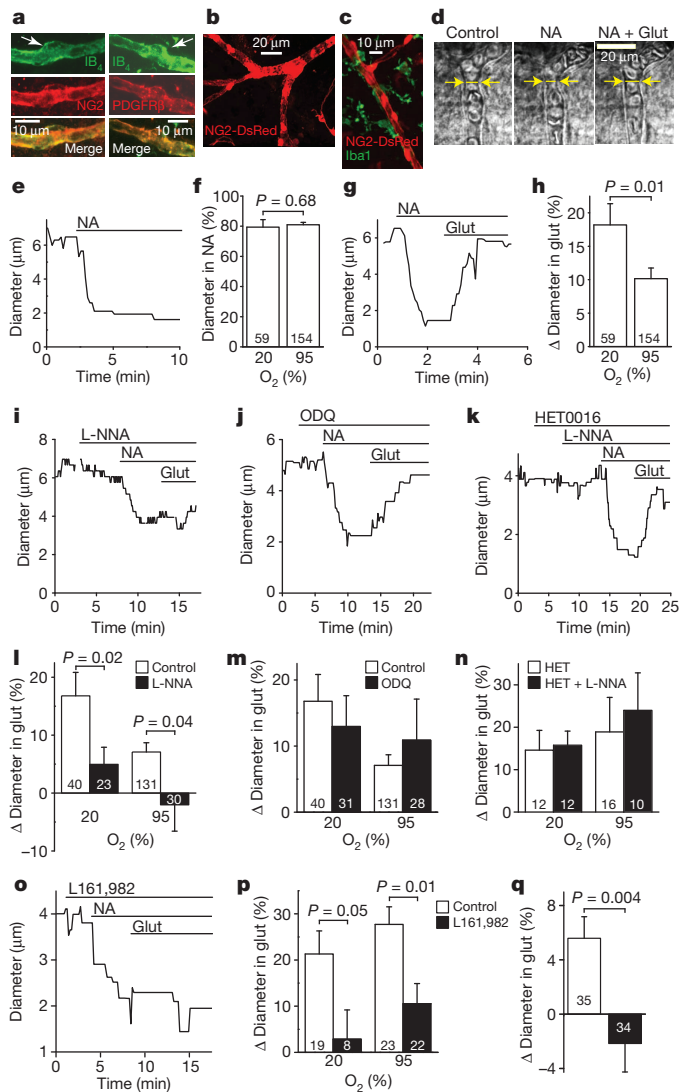


Figure 1 | Signalling pathways that control capillary diameter. **a**, Capillaries in the molecular layer of rat cerebellum labelled using isolectin B₄; pericytes (arrows) labelled with NG2 or PDGFR β antibodies. **b**, Cerebellar capillaries in NG2-DsRed mouse; pericytes are red. **c**, Neocortical capillaries in NG2-DsRed mouse; microglia labelled for Iba1. **d**, Rat cerebellar capillary response to 2 μ M noradrenaline (NA) and superimposed 500 μ M glutamate (Glut). Line shows lumen diameter. **e**, NA-evoked prolonged constriction (95% O₂; a large constriction is shown for clarity). **f**, Diameter in NA was not affected by [O₂] (graphs show percentage of baseline diameter before drugs). **g**, Glutamate dilates capillaries (20% O₂; a large dilation is shown for clarity). **h**, Dilation was larger in low [O₂]. **i**, NOS blocker L-NNA (100 μ M) inhibits Glut-evoked dilation (20% O₂). **j**, Guanylyl cyclase blocker ODQ (10 μ M) does not block dilation (20% O₂). **k**, Blocking 20-HETE production (HET0016, 1 μ M) abolishes the inhibitory effect of L-NNA (20% O₂). **l**, L-NNA reduces glutamate-evoked dilations at high and low [O₂] ($P = 0.002$, ANOVA on all data; P values at each [O₂] from post-hoc t -tests). **m**, ODQ does not affect glutamate-evoked dilation. **n**, HET0016 abolishes effect of L-NNA. **o**, **p**, Blocking EP₄ receptors (L161,982, 1 μ M) inhibits glutamate-evoked dilation (**o**, 20% O₂) at high and low [O₂] (**p**). Data in **d–p** are from rat cerebellar capillaries. **q**, EP₄ block abolishes glutamate-evoked dilation in rat neocortical capillaries (20% O₂). Drug effects on baseline diameter are in Extended Data Fig. 2. Data shown as mean \pm s.e.m.

variance (ANOVA); L-NNA and other blockers used did not inhibit the noradrenaline-evoked constriction, see Extended Data Fig. 2). Unexpectedly, the dilation was not affected by blocking guanylyl cyclase with ODQ (1*H*-[1,2,4]Oxadiazolo[4,3-*a*]quinoxalin-1-one, 10 μ M, $P = 1$, ANOVA), so NO does not act by raising [cGMP] in the pericyte (Fig. 1j, m

and Extended Data Fig. 2e). However, when production of vasoconstricting 20-HETE was blocked using HET0016 (1 μ M), L-NNA no longer inhibited the dilation (Fig. 1k, n; $P = 0.0005$ from ANOVA on black bars in **l** and **n**), implying that NO promotes dilation by preventing 20-HETE formation. As a robust dilation occurs with both NO and 20-HETE synthesis blocked (Fig. 1k, n), another messenger must be active. Blocking synthesis of epoxy derivatives of arachidonic acid with the inhibitor MS-PPH (10 μ M) did not affect the dilation (Extended Data Fig. 2i, ANOVA $P = 0.92$), but blocking EP₄ receptors for prostaglandin E₂ (with 1 μ M L-161,982) greatly reduced it (Fig. 1o–p, $P = 0.001$, ANOVA). A similar inhibition of capillary dilation by blocking EP₄ receptors was seen in neocortical pericytes (Fig. 1q, $P = 0.004$). Application of prostaglandin E₂ (PGE₂) itself dilated cerebellar capillaries (Extended Data Fig. 2n). We therefore identify the messenger that dilates capillaries in response to glutamate as prostaglandin E₂ (or a related species active at EP₄ receptors), but this dilation requires NO release to suppress 20-HETE formation (Extended Data Fig. 1b).

Glutamate (500 μ M) or NMDA (100 μ M) (Fig. 2a, b, d), or parallel fibre stimulation (Fig. 2c, d), evoked an outward membrane current in pericytes patch-clamped at -55 to -75 mV (sometimes preceded by a small inward current; Extended Data Fig. 3). The stimulation-evoked outward current was inhibited ($P = 0.005$, paired t -test) by blocking action potentials with tetrodotoxin (TTX) (Fig. 2c, reduced to $15 \pm 12\%$ ($n = 4$) of its amplitude without TTX, not significantly different from zero, $P = 0.28$), whereas the NMDA-evoked current was unaffected by TTX (reduced by $12 \pm 19\%$, $n = 5$, $P = 0.56$), consistent with stimulation evoking the outward current by generating action potentials that release glutamate. The stimulation-evoked current of approximately 30 pA is expected to hyperpolarize the cells by approximately 9 mV (see Methods) and decrease voltage-gated Ca²⁺ entry, causing active relaxation⁹ (although other Ca²⁺ sources, and cyclic nucleotides, may also regulate contractile tone^{9,20}). Indeed, parallel fibre stimulation produced a dilation of $14.9 \pm 3.1\%$ in 21 capillaries in 20% O₂ (Fig. 2e–h and Supplementary Video 2), which (unlike the noradrenaline-evoked constriction) was blocked by TTX and by blocking EP₄ receptors (Fig. 2g, h).

An outward current at negative potentials is not consistent with activation of glutamatergic ionotropic receptors, but is consistent with K⁺ current activation. These data suggest that glutamate release generates PGE₂, which dilates the capillaries by activating an outward K⁺ current in pericytes. PGE₂ similarly activates an outward K⁺ current in aortic smooth muscle²¹, and relaxes kidney pericytes²².

Pericytes increase blood flow *in vivo*

To assess whether pericyte relaxation regulates blood flow *in vivo*, we electrically stimulated the whisker pad (at 3 Hz) and used two-photon imaging of the vasculature (labelled with FITC-dextran) in somatosensory cortex to monitor dilations of penetrating arterioles (entering the cortex from the pia) and capillaries, in anaesthetized mice expressing DsRed in pericytes (Fig. 3a). The capillary diameter *in vivo* was 4.4 ± 0.1 μ m in 633 capillary regions (Extended Data Table 1). Pericytes were visualized up to 200 μ m deep in the cortex (layer 2/3). Brief whisker pad stimulation (2 s) evoked vessel dilations that peaked just after the end of the stimulation (~ 2.5 s, Fig. 3b). Longer (15 s) stimulation produced dilations that initially followed the same time course, then dilated further throughout the stimulation (Fig. 3b and Extended Data Fig. 4). Most imaging employed 15 s stimuli, which increased the response magnitude and measurement accuracy. Repeated stimulation gave reproducible responses (Extended Data Fig. 5a, b). We segmented the vasculature by the branching order of the vessels, zero being the penetrating arteriole, one being the primary capillary branching off the arteriole, and so on (Extended Data Fig. 1; see Extended Data Table 1 for resting diameters, dilations and numbers of each vessel order). Whisker pad stimulation dilated vessels of all orders (Fig. 3c and Extended Data Table 1). The fraction of vessels responding (that is, with a dilation $> 5\%$) was similar in penetrating arterioles and in first-order capillaries, whereas

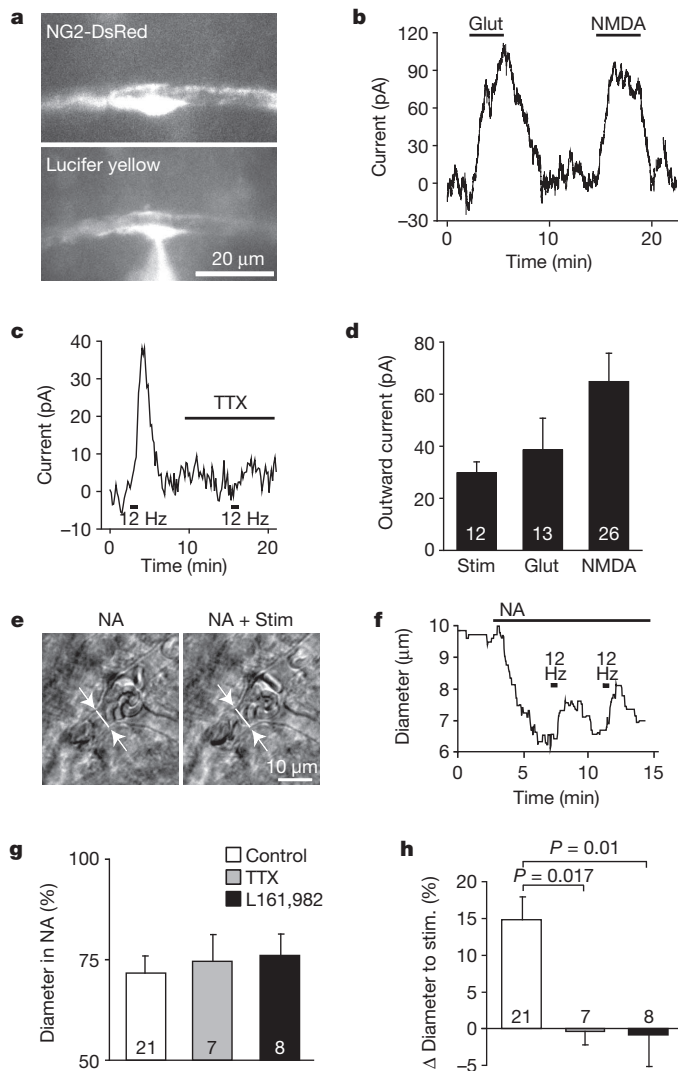


Figure 2 | Pericyte membrane current and capillary dilation in cerebellar slices. **a**, DsRed-labelled patch-clamped pericyte in molecular layer of mouse cerebellum. Lucifer yellow from pipette overlaps with DsRed. **b**, Glutamate (Glut, 500 μ M) and NMDA (100 μ M) evoked outward current (\sim 55 mV). **c**, Parallel fibre stimulation (Stim, 12 Hz)-evoked outward current (\sim 74 mV) is blocked by 1 μ M TTX. **d**, Mean outward currents evoked by stimulation, glutamate and NMDA (**b–d** are in 95% O_2). **e**, **f**, Parallel fibre stimulation (in NA 1 μ M) in rat cerebellar slice evokes capillary dilation (20% O_2). **g**, **h**, Constriction by NA (**g**) was unaffected by TTX or EP₄ block (L161,982), which abolished the stimulation-evoked dilation (**h**). *P* values from one-way ANOVA with Dunnett's post-hoc tests. NA, noradrenaline. Data shown as mean \pm s.e.m.

the frequency of capillary responses decreased with increasing order (Fig. 3c).

To establish where vasodilation is initiated, we imaged different orders of vessel simultaneously. Strikingly, first-order capillaries usually dilated before penetrating arterioles (Fig. 3d, e and Supplementary Video 3), with vasodilation onset (assessed as the time to 10% of the maximum dilation) in the capillary being 1.38 ± 0.38 s earlier than for the penetrating arteriole (Fig. 3e, f, $P = 0.015$). Further along the vascular tree there was no significant difference in the time to dilation of simultaneously imaged capillaries of adjacent order (Fig. 3f and Extended Data Fig. 5c). Thus, capillaries dilate before the penetrating arteriole feeding them. Averaging over all vessels of the same order (not just those imaged simultaneously) showed a similar faster dilation of capillaries than of penetrating arterioles (Fig. 3g), with the time to 10% of the maximum dilation for penetrating arterioles (3.7 ± 0.3 s) being significantly longer

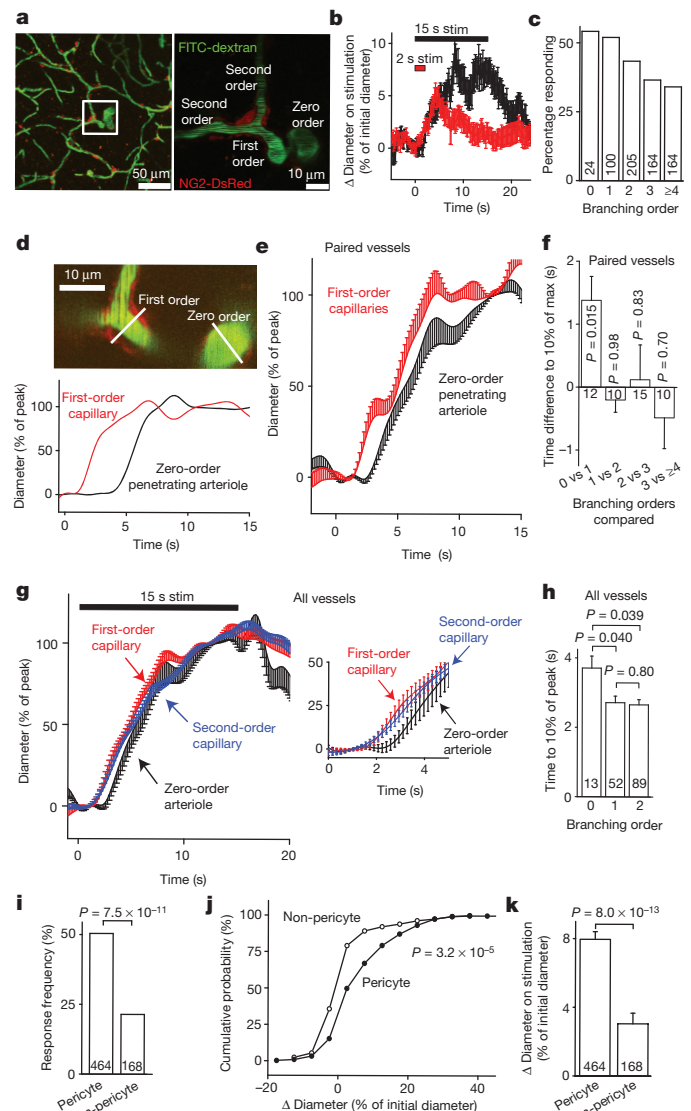


Figure 3 | Active dilation of capillaries by pericytes *in vivo* in mouse cerebellar cortex. **a**, Confocal stack (90 μ m thick, maximum intensity projection) of FITC-dextran-filled vessels *in vivo* in somatosensory cortex of NG2-DsRed mouse (pericytes are red). Enlargement (single image) shows a penetrating arteriole (0th order) giving off a capillary (first order) which splits into second-order branches. **b**, Response of 45 capillary regions to 2-s and 15-s whisker-pad stimulation. **c**, Percentage of vessel regions of different orders (number studied on bars) showing $>5\%$ dilation to stimulation. **d**, Simultaneous imaging (top, lines show measurement loci) of penetrating arteriole and first-order capillary: capillary dilates 3 s before arteriole (bottom, smoothing in **d–g** is explained in Methods and Extended Data Fig. 4). **e**, Dilation time course in simultaneously imaged penetrating arterioles and first-order capillaries. **f**, Time to 10% of peak dilation for (*j–l*)th order (third order for $j \geq 4$) vessel minus that of *j*th-order vessel (*j* is an integer ≥ 1). Capillaries dilate faster than arterioles. **g**, Dilation time course in all responding ($>5\%$) penetrating arterioles and first- and second-order capillaries. Inset expands initial response. **h**, Time to 10% of peak dilation in all zero- to second-order responding vessels. **i**, Percentage of capillary locations with or without pericytes showing $>5\%$ dilations. **j**, Cumulative probability of capillary diameter changes (including 'non-responding' capillaries with $<5\%$ dilations) in 464 pericyte and 168 non-pericyte locations. Diameter changes $<0\%$ (constrictions) represent random changes and measurement error. **k**, Mean responses for distributions in **j** (*P* values from Mann–Whitney U-tests). Data shown as mean \pm s.e.m.

than the values (~ 2.7 s) obtained for first- and second-order capillaries ($P = 0.040$ and 0.039 respectively, Fig. 3h and Extended Data Fig. 5d). As expected, the time course of the blood flow increase in capillaries,

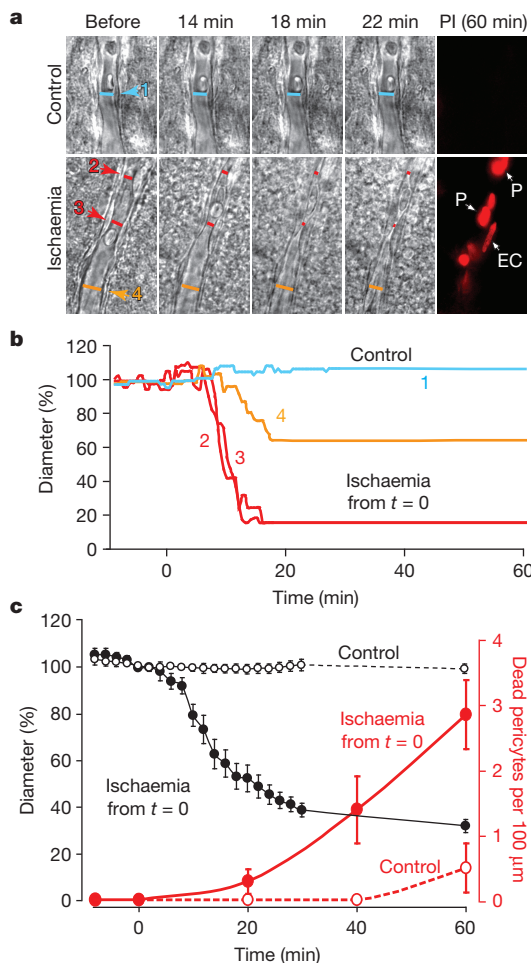


Figure 4 | In ischaemia, pericytes constrict capillaries and then die in rigor.

a, Top, capillary in a P21 rat cortical slice in normal solution. Bottom, capillary exposed to simulated ischaemia. Right, propidium iodide (PI, 7.5 μM) labelling after one hour, showing dead pericytes (P) and an endothelial cell (EC). **b**, Vessel diameters at regions indicated in **a**, over time (red lines are for diameter near pericytes that die). **c**, Mean diameter and number of dead pericytes per 100 μm from nine capillaries in ischaemia (measured at 18 locations) and six capillaries in normal solution (13 locations). Diameter in ischaemia was reduced from control ($P = 1.3 \times 10^{-15}$ and 7.7×10^{-17} at 30 and 60 min). Pericyte death was higher in ischaemia at 40 and 60 min ($P = 0.041$ and 0.021 , Mann–Whitney U-test). A few pericytes also died on capillaries in non-ischaemic slices, but did not constrict capillaries ($1.3 \pm 1.5\%$ diameter decrease at three dead pericytes on two vessels). Data shown as mean \pm s.e.m.

assessed from the speed of red blood cell movement with line-scanning²³, was similar to that of the capillary dilation (Extended Data Fig. 5f).

The faster dilation in capillaries compared to arterioles indicates that capillary dilation is not a passive response to a pressure increase produced by arteriole dilation. To assess whether pericytes generate this dilation, we measured the diameter changes of capillaries at locations where DsRed-labelled pericytes were present (either somata or processes, responses did not differ significantly at these locations; Extended Data Fig. 5e) or where no pericyte was visible. The resting diameter of capillaries was larger where pericyte somata or processes were present ($4.62 \pm 0.09 \mu\text{m}$, $n = 464$) than in pericyte-free zones ($3.72 \pm 0.08 \mu\text{m}$, $n = 168$, Mann–Whitney $P = 2.7 \times 10^{-7}$), suggesting that pericytes induce an increase of capillary diameter. Dilations over 5% were much more frequent at pericyte locations (Fig. 3i; $P = 7.5 \times 10^{-11}$, chi-squared test), where the responses were also larger ($P = 3.2 \times 10^{-5}$, Kolmogorov–Smirnov test; Fig. 3j, k). These data confirm that pericytes actively relax to generate the capillary dilation.

Pericytes constrict and die in ischaemia

A key question is whether pericyte control of capillary diameter also has a role in pathology. Pericytes constrict some retinal capillaries in ischaemia¹, perhaps because pericyte $[\text{Ca}^{2+}]_i$ rises when ion pumping is inhibited by ATP depletion. Cortical capillaries also constrict following middle cerebral artery occlusion¹⁴ (MCAO) *in vivo*. Clinically, reperfusion of thrombus-occluded arteries using tissue plasminogen activator can be achieved in 1 to 6 h^{24,25} but, even when arterial flow is restored, a long-lasting reduction of cerebral blood flow can ensue^{26–29}. This may reflect pericyte constriction outlasting ischaemia¹⁴, but it is unclear why the constriction is so prolonged. We examined the effect of ischaemia on pericyte health using propidium iodide to mark cell death.

Live imaging of cerebral cortical slices exposed to simulated ischaemia (oxygen-glucose deprivation with ATP synthesis by glycolysis and oxidative phosphorylation inhibited with iodoacetate and antimycin) revealed that, within approximately 15 min, grey matter capillaries constricted at spatially restricted regions near pericytes (Fig. 4a–c). In contrast, the diameter of capillaries not exposed to ischaemia was stable over 60 min (reduced by $3.2 \pm 3.0\%$, $P = 0.31$, $n = 13$; Fig. 4c). Although capillaries that were not exposed to ischaemia showed little pericyte death even after 1 hour (as assessed by propidium iodide labelling), ischaemia led to most pericytes on capillaries dying after approximately 40 min, usually at locations at which the earlier constriction had occurred. All capillaries exposed to ischaemia that we examined showed a consistent response, in which the pericytes first constricted the capillaries, and then died (Fig. 4c). Death of pericytes in rigor, after they have been constricted by a loss of energy supply, should produce a long-lasting increase in the resistance of the capillary bed.

To sample more pericytes than is possible while live imaging the capillary diameter, and to examine mechanisms contributing to their death, we acquired confocal stacks of brain slices exposed to simulated ischaemia, which were then fixed and labelled for NG2 and/or isolectin B₄. Ischaemia led to pericytes apposed to capillaries dying rapidly in both the white matter of the cerebellum (Fig. 5a, b) and the grey matter of the cerebral cortex (Fig. 5c, d). For pericytes exposed to simulated ischaemia as above, approximately 90% of pericytes died within 1 hour (Fig. 5b). This was unaffected by blocking action potentials (with TTX, 1 μM) but was halved by blocking AMPA (α -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid)/kainate (25 μM NBQX) or NMDA receptors (50 μM D-AP5, 50 μM MK-801 and 100 μM 7-chlorokynureate), implying an excitotoxic contribution to pericyte death. When oxygen-glucose deprivation (OGD) alone was used (without antimycin and iodoacetate), to allow ATP generation when oxygen and glucose were restored, approximately 40% of pericytes died after 1 hour, but OGD-evoked death increased 1.5-fold during 1 h of reperfusion (Fig. 5c, d). Ionotropic glutamate receptor block or removal of external Ca^{2+} again significantly reduced the death (Fig. 5d), while blocking NO production had a small protective effect and lowering free radical levels by scavenging O_2^- had no significant effect (Fig. 5d and Extended Data Fig. 6a). Blocking metabotropic glutamate receptors or 20-HETE production, which might prevent $[\text{Ca}^{2+}]_i$ rises, or blocking mitochondrial calcium uptake, also had no effect (Extended Data Fig. 6b).

Pericytes apposed to capillaries, unlike endothelial cells, also died *in vivo* after 90 min of MCAO (followed by 22.5 h recovery; Fig. 5e, f, and Extended Data Fig. 6c). In contrast, a sham operation occluding only the internal carotid artery (see Methods) produced less pericyte death, and a sham operation with no artery occlusion (which did not affect blood flow; see Methods) induced no more death than in naive untreated animals. Thus, pericyte death is a rapid response of the cerebral vascular bed to ischaemia, both in brain slices and *in vivo*.

Discussion

Understanding what initiates the blood-flow increase in response to neuronal activity is crucial for understanding both how information processing is powered and how functional imaging signals are generated³⁰. Most neurons are closer to capillaries ($\sim 8.4 \mu\text{m}$ distant, in hippocampus³¹)

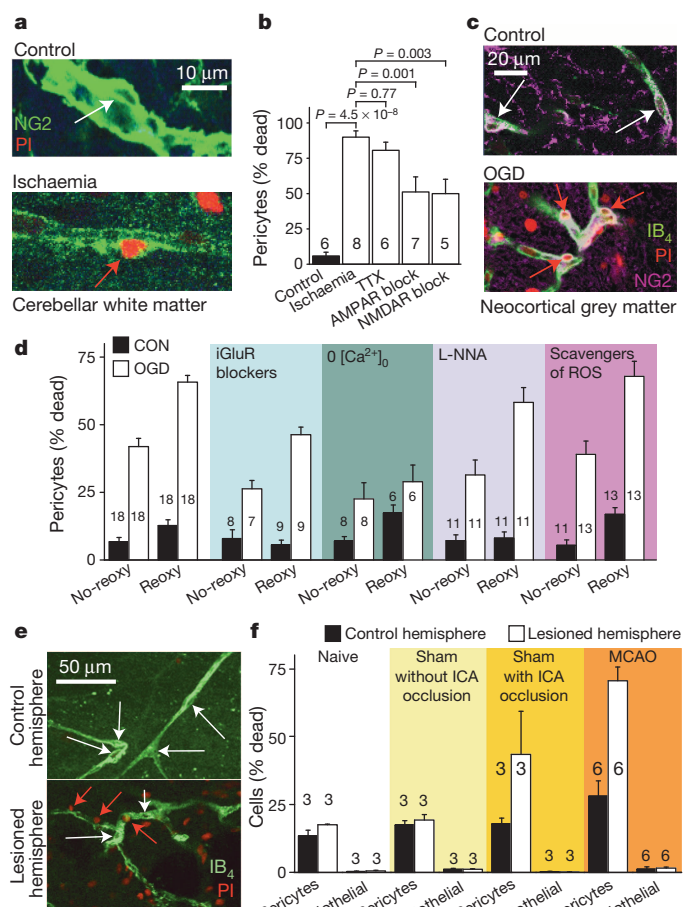


Figure 5 | Pericyte death in ischaemia. **a**, Rat cerebellar slice white matter capillaries labelled for NG2 and propidium iodide (PI) after 1 h of control (white arrow, living pericyte) or ischaemia solution containing antimycin and iodoacetate (red arrow, dead pericyte). **b**, Percentage of pericytes dead in control or after 1 h ischaemia (as in **a**) alone or with block of action potentials (TTX, 1 μ M), AMPA/kainate receptors (25 μ M NBQX), or NMDA receptors (50 μ M D-AP5, 50 μ M MK-801, 100 μ M 7-chlorokynurenenate); *P* values from one-way ANOVA with Dunnett's post-hoc tests. **c**, Rat neocortical slice grey matter capillaries labelled for IB₄, NG2 and PI after 1 h of control solution or oxygen and glucose deprivation (OGD). **d**, Percentage of pericytes (as in **c**) dead after 1 h OGD (No-reoxy) or OGD followed by 1 h of control solution (Reoxy) with no drugs, or with iGluR block (NBQX (25 μ M), AP5 (50 μ M) and 7CK (100 μ M)), zero $[Ca^{2+}]_o$, NOS block (100 μ M L-NNA), or free radical scavenging (150 μ M MnTBAP or 100 μ M PBN, pooled data from Extended Data Fig. 5a) throughout. OGD killed pericytes ($P = 10^{-13}$, ANOVA) and death increased during reperfusion ($P = 3.3 \times 10^{-13}$). iGluR block or zero $[Ca^{2+}]_o$ reduced death ($P = 2.7 \times 10^{-4}$ and 6.0×10^{-7} , ANOVA with Dunnett's post-hoc test). Blocking NOS had a small protective effect ($P = 0.026$); reactive oxygen species (ROS) scavenging did not ($P = 0.99$). **e**, **f**, Confocal images of striatal capillaries labelled with IB₄ and PI (**e**) and percentage of striatal pericytes and endothelial cells that are dead (**f**) from the control and lesioned hemisphere of *in vivo* MCAO-treated rats (90 min, assessed 22.5 h later), sham-operated rats (with or without filament being inserted into the internal carotid artery (ICA)), and naive control animals. More pericytes die than endothelial cells ($P = 10^{-6}$, repeated measures ANOVA). For pericytes, but not endothelial cells, cell death is greater in lesioned hemisphere (main effect of hemisphere, $P = 0.004$; hemisphere-cell type interaction $P = 0.003$) and cell death is greater in MCAO-lesioned animals than in naive or sham animals without ICA occlusion ($P = 0.005$ and 0.01 , Tukey post-hoc tests). See Extended Data Fig. 5 for data from cortex. Data shown as mean \pm s.e.m.

than to arterioles (70 μ m distant³¹), suggesting that neurons might adjust their energy supply by initially signalling to pericytes (Extended Data Fig. 1a). Our data support this concept: neuronal activity leads to a release of messengers that activate an outward membrane current in pericytes (Fig. 2) and dilate capillaries before arterioles (Fig. 3). Capillary dilation

implies that there is a resting tone set by the pericytes, perhaps as a result of noradrenaline release from locus coeruleus axons, two-thirds of the perivascular terminals of which end near capillaries rather than arterioles³². Vascular diameter responses can propagate between adjacent pericytes^{1,9}, but it is not yet known whether arterioles receive a signal to dilate from pericytes, or from vasoactive messengers which reach arterioles later than they reach pericytes. We have identified the cause of pericyte dilation as being EP₄ receptor activation, by PGE₂ or a related compound, although NO production is also needed to suppress synthesis of vasoconstricting 20-HETE (Fig. 1 and Extended Data Fig. 1b). These mechanisms resemble those controlling arteriole dilation, and may reflect glutamate release activating the production of arachidonic acid and its derivatives in astrocytes or neurons¹³.

The more frequent occurrence of capillary dilation at pericyte locations (Fig. 3i–k), and the faster onset of capillary dilation than of arteriole dilation *in vivo* (Fig. 3d–h) suggest that pericytes actively relax to dilate capillaries. Our data suggest that capillaries have two conceptually separate roles in regulating cerebral blood flow. First, being closer to neurons, they detect neuronal activity earlier than arterioles can, and may pass a hyperpolarizing vasodilatory signal back to arterioles (through gap junctions between pericytes or endothelial cells)^{1,9}. Second, capillary vasodilation itself contributes significantly to increasing blood flow. To assess the extent to which capillary dilations increase blood flow, we used data on the vascular tree in mouse cortex³³ (see Methods). For a mean arteriole dilation of 5.9% during prolonged (15 s) stimulation (Extended Data Table 1), a capillary dilation of 6.7% (averaged over all capillary orders, Extended Data Table 1) and ignoring venule dilation, the steady state blood flow was predicted to increase by 19%. Omitting the capillary dilation predicted a flow increase of only 3%. Thus, capillary dilation is estimated to generate $100 \times (19 - 3)/19 = 84\%$ of the steady-state increase in blood flow evoked by neuronal activity, and capillaries dilate approximately 1 s before penetrating arterioles (Fig. 3d–h). These results suggest that BOLD functional imaging signals largely reflect capillary dilation by pericytes.

For ischaemia, our data support, but modify, the suggestion that pericyte constriction^{1,14} may be a cause of the long-lasting decrease of cerebral blood flow that occurs even when a blocked artery is opened up after stroke^{26–29}. It was previously envisaged that constriction of capillaries was by healthy pericytes and could be reversed by suppressing oxidative stress¹⁴, whereas we find that, after ischaemia has constricted them (Fig. 4), pericytes die readily (Fig. 5). This death is mediated in part by glutamate, but is not reduced by free radical scavenging, suggesting that the constriction and death differ at least partly in their causes. Pericyte death in rigor will produce a long-lasting decrease of capillary blood flow^{26–29}, as well as a breakdown of the (pericyte-maintained^{4–6}) blood–brain barrier. Both of these will contribute to ongoing neuronal damage, highlighting the potential importance of preventing pericyte death as a therapeutic strategy after stroke, particularly in the penumbra of an affected region. To develop this approach, it will be necessary to develop small molecule inhibitors of pericyte death that could be administered, perhaps with tissue plasminogen activator, soon after a stroke has occurred.

METHODS SUMMARY

Brain slices from rats or NG2-DsRed mice were made and capillaries were imaged as described previously¹. The fluorescence of DsRed-labelled capillary pericytes was used to identify pericytes for patch-clamping. Two-photon imaging of FITC-dextran-labelled cortical vessels and DsRed-labelled pericytes *in vivo* was performed in mice as previously described for cerebellum³⁴.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 29 July 2013; accepted 19 February 2014.

Published online 26 March 2014.

1. Peppiatt, C. M., Howarth, C., Mobbs, P. & Attwell, D. Bidirectional control of CNS capillary diameter by pericytes. *Nature* **443**, 700–704 (2006).

2. Fernández-Klett, F., Offenhauser, N., Dirnagl, U., Priller, J. & Lindauer, U. Pericytes in capillaries are contractile *in vivo*, but arterioles mediate functional hyperemia in the mouse brain. *Proc. Natl. Acad. Sci. USA* **107**, 22290–22295 (2010).
3. Chan-Ling, T. *et al.* Desmin ensheathment ratio as an indicator of vessel stability: evidence in normal development and in retinopathy of prematurity. *Am. J. Pathol.* **165**, 1301–1313 (2004).
4. Bell, R. D. *et al.* Pericytes control key neurovascular functions and neuronal phenotype in the adult brain and during brain aging. *Neuron* **68**, 409–427 (2010).
5. Armulik, A. *et al.* Pericytes regulate the blood–brain barrier. *Nature* **468**, 557–561 (2010).
6. Daneman, R., Zhou, L., Kebede, A. A. & Barres, B. A. Pericytes are required for blood–brain barrier integrity during embryogenesis. *Nature* **468**, 562–566 (2010).
7. Göritz, C. *et al.* A pericyte origin of spinal cord scar tissue. *Science* **333**, 238–242 (2011).
8. Dore-Duffy, P., Katychew, A., Wang, X. & Van Buren, E. CNS microvascular pericytes exhibit multipotential stem cell activity. *J. Cereb. Blood Flow Metab.* **26**, 613–624 (2006).
9. Puro, D. G. Physiology and pathobiology of the pericyte-containing retinal microvasculature: new developments. *Microcirculation* **14**, 1–10 (2007).
10. Villringer, A., Them, A., Lindauer, U., Einhäupl, K. & Dirnagl, U. Capillary perfusion of the rat brain cortex. An *in vivo* confocal microscopy study. *Circ. Res.* **75**, 55–62 (1994).
11. Schulte, M. L., Wood, J. D. & Hudetz, A. G. Cortical electrical stimulation alters erythrocyte perfusion pattern in the cerebral capillary network of the rat. *Brain Res.* **963**, 81–92 (2003).
12. Jespersen, S. N. & Østergaard, L. The roles of cerebral blood flow, capillary transit time heterogeneity, and oxygen tension in brain oxygenation and metabolism. *J. Cereb. Blood Flow Metab.* **32**, 264–277 (2012).
13. Attwell, D. *et al.* Glial and neuronal control of blood flow. *Nature* **468**, 232–243 (2010).
14. Yemisci, M. *et al.* Pericyte contraction induced by oxidative-nitrative stress impairs capillary reflow despite successful opening of an occluded cerebral artery. *Nature Med.* **15**, 1031–1037 (2009).
15. Vates, G. E., Takano, T., Zlokovic, B. & Nedergaard, M. Pericyte constriction after stroke: the jury is still out. *Nature Med.* **16**, 959 (2010).
16. Hall, C. N., Klein-Flügge, M. C., Howarth, C. & Attwell, D. Oxidative phosphorylation, not glycolysis, powers presynaptic and postsynaptic mechanisms underlying brain information processing. *J. Neurosci.* **32**, 8940–8951 (2012).
17. Hall, C. N. & Attwell, D. Assessing the physiological concentration and targets of nitric oxide in brain tissue. *J. Physiol. (Lond.)* **586**, 3597–3615 (2008).
18. Zhu, X., Bergles, D. E. & Nishiyama, A. NG2 cells generate both oligodendrocytes and gray matter astrocytes. *Development* **135**, 145–157 (2008).
19. Krueger, M. & Bechmann, I. CNS pericytes: concepts, misconceptions, and a way out. *Glia* **58**, 1–10 (2010).
20. Hamilton, N. B., Attwell, D. & Hall, C. N. Pericyte-mediated regulation of capillary diameter: a component of neurovascular coupling in health and disease. *Front. Neuroenergetics* **2**, 5 (2010).
21. Serebryakov, V., Zakharenko, S., Snetkov, V. & Takeda, K. Effects of prostaglandins E₁ and E₂ on cultured smooth muscle cells and strips of rat aorta. *Prostaglandins* **47**, 353–365 (1994).
22. Crawford, C. *et al.* An intact kidney slice model to investigate vasa recta properties and function *in situ*. *Nephron Physiol.* **120**, 17–31 (2012).
23. Kleinfeld, D., Mitra, P. P., Helmchen, F. & Denk, W. Fluctuations and stimulus-induced changes in blood flow observed in individual capillaries in layers 2 through 4 of rat neocortex. *Proc. Natl. Acad. Sci. USA* **95**, 15741–15746 (1998).
24. Khatri, P. *et al.* Good clinical outcome after ischemic stroke with successful revascularization is time-dependent. *Neurology* **73**, 1066–1072 (2009).
25. Gupta, E. *et al.* Higher volume endovascular stroke centers have faster times to treatment, higher reperfusion rates and higher rates of good clinical outcomes. *J. Neurointerv. Surg.* **5**, 294–297 (2013).
26. Hauck, E. F., Apostel, S., Hoffmann, J. F., Heimann, A. & Kempski, O. Capillary flow and diameter changes during reperfusion after global cerebral ischemia studied by intravital video microscopy. *J. Cereb. Blood Flow Metab.* **24**, 383–391 (2004).
27. Leffler, C. W., Beasley, D. G. & Busija, D. W. Cerebral ischemia alters cerebral microvascular reactivity in newborn pigs. *Am. J. Physiol.* **257**, H266–H271 (1989).
28. Nelson, C. W., Wei, E. P., Povlishock, J. T., Kontos, H. A. & Moskowitz, M. A. Oxygen radicals in cerebral ischemia. *Am. J. Physiol.* **263**, H1356–H1362 (1992).
29. Baird, A. E. *et al.* Reperfusion after thrombolytic therapy in ischemic stroke measured by single-photon emission computed tomography. *Stroke* **25**, 79–85 (1994).
30. Attwell, D. & Iadecola, C. The neural basis of functional imaging signals. *Trends Neurosci.* **25**, 621–625 (2002).
31. Lovick, T. A., Brown, L. A. & Kay, B. J. Neurovascular relationships in hippocampal slices: physiological and anatomical studies of mechanisms underlying flow-metabolism coupling in intraparenchymal microvessels. *Neuroscience* **92**, 47–60 (1999).
32. Cohen, Z., Molinatti, G. & Hamel, E. Astroglial and vascular interactions of noradrenaline terminals in the rat cerebral cortex. *J. Cereb. Blood Flow Metab.* **17**, 894–904 (1997).
33. Blinder, P. *et al.* The cortical angiome: a 3-D interconnected vascular network with noncolumnar patterns of blood flow. *Nature Neurosci.* **16**, 889–897 (2013).
34. Mathiesen, C., Brazhe, A., Thomsen, K. & Lauritzen, M. Spontaneous calcium waves in Bergmann glia increase with age and hypoxia and may reduce tissue oxygen. *J. Cereb. Blood Flow Metab.* **33**, 161–169 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank B. Clark, A. Gibb, A. Gourine, C. Howarth, R. Jolivet, C. Madry, P. Mobbs, B. Richardson and A. Silver for comments on the manuscript. This work was supported by the Fondation Leducq, European Research Council, Wellcome Trust, UK Medical Research Council, Rosetrees Trust, Nordea Foundation via the Center for Healthy Aging, the Lundbeck Foundation, NOVO-Nordisk Foundation and Danish Medical Research Council.

Author Contributions All authors (C.N.H., C.R., B.G., N.B.H., A.M., B.A.S., F.M.O., A.M.B., M.L. and D.A.) contributed to designing the study, doing the experiments, analysing the results and writing the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.A. (d.attwell@ucl.ac.uk) or M.L. (mlauritz@sund.ku.dk).

Mechanism of Tc toxin action revealed in molecular detail

Dominic Meusch^{1*}, Christos Gatsogiannis^{1*}, Rouslan G. Efremov^{1*}, Alexander E. Lang², Oliver Hofnagel¹, Ingrid R. Vetter³, Klaus Aktories^{2,4} & Stefan Raunser^{1,5}

Tripartite Tc toxin complexes of bacterial pathogens perforate the host membrane and translocate toxic enzymes into the host cell, including in humans. The underlying mechanism is complex but poorly understood. Here we report the first, to our knowledge, high-resolution structures of a TcA subunit in its prepore and pore state and of a complete 1.7 megadalton Tc complex. The structures reveal that, in addition to a translocation channel, TcA forms four receptor-binding sites and a neuraminidase-like region, which are important for its host specificity. pH-induced opening of the shell releases an entropic spring that drives the injection of the TcA channel into the membrane. Binding of TcB/TcC to TcA opens a gate formed by a six-bladed β -propeller and results in a continuous protein translocation channel, whose architecture and properties suggest a novel mode of protein unfolding and translocation. Our results allow us to understand key steps of infections involving Tc toxins at the molecular level.

Tc toxin complexes are virulence factors of many bacteria such as the plague pathogen *Yersinia pestis* and the insect pathogen *Photorhabdus luminescens*^{1–3}. Excreted by the bacteria as soluble proteins, Tcs bind to the cell surface, are endocytosed and perforate the host endosomal membrane by forming channels that translocate toxic enzymes into the host^{4,5}. This damages and ultimately kills the target cells. Tcs with specificity for insects are receiving great interest as potential biopesticides for expression in transgenic plants as alternatives to *Bacillus thuringiensis* toxins^{6,7}. Tc toxins are composed of TcA, TcB and TcC subunits^{4,8}. TcB and TcC together form a closed cage, in which the cytotoxic domain of TcC is autoproteolytically cleaved⁹. TcA forms a large bell-shaped pentameric structure^{5,10,11} and enters the membrane like a syringe, forming a translocation channel through which the cytotoxic domain is probably transported into the cytoplasm⁵. This unique mechanism differs from that of typical pore-forming toxins¹² or other toxins that form translocation pores, for example, diphtheria¹³ and anthrax toxin¹⁴. However, little is known about the force that drives this mechanism and how the three components interact with each other and with the host membrane to unfold and translocate the toxic domain. A greater understanding requires knowledge of the structures of the Tc complex and its subunits in different states, that is, before and after membrane insertion. We use a hybrid approach combining X-ray crystallography and electron cryomicroscopy (cryo-EM) to determine the crystal structures of the prepore TcA (TcdA1) subunit and TcB–TcC (TcdB2–TccC3) subunits, the cryo-EM structures of TcA in its pore state and the complete Tc toxin complex PTC3.

TcA has four receptor-binding domains

To understand how the pentameric TcA subunit (TcdA1 from *P. luminescens*) is built in molecular detail and how it functions, we solved its crystal structure—comprising 12,500 residues with a molecular weight of 1.41 MDa—at a resolution of 4.0 Å (Fig. 1a and Supplementary Video 1). The TcA protomer is composed of eight domains (Fig. 1b, c). Six domains form the shell and are connected by a ~42 amino acid linker to two domains forming the inner channel (Fig. 1b). The pentamer is

formed through the tight assembly of the protomers, which interact with all four other protomers (Extended Data Table 1a).

The large extended α -helical shell domain (amino acids 14–297, 434–1090, 1608–1632, 1762–1972) forms the inner scaffold of the shell (Fig. 1b). It contains over 60 irregularly arranged α -helices, including many short helices, and no β -strands, and can be separated into a larger and smaller lobe that are arranged perpendicular to each other (Fig. 1b). We could not identify any structural relative, suggesting that this structure represents a previously unknown fold type.

The α -helical shell domain contains three large insertions, forming a neuraminidase-like domain and four receptor-binding domains (Fig. 1b, c). The neuraminidase-like domain (1091–1307, 1581–1607), which is structurally homologous to neuraminidases¹⁵ (Extended Data Fig. 1), closes the bottom of the pentameric shell. The receptor-binding domains A–D (A, 298–433; B, 1308–1382, 1491–1580; C, 1383–1490; and D, 1633–1761) have an immunoglobulin (Ig)-like β -sandwich fold of two sheets with antiparallel β -strands (Extended Data Fig. 1). The domains are structurally reminiscent of the receptor-binding domains of the diphtheria¹⁶ and anthrax toxins¹⁷, suggesting that these domains act as receptor-binding domains. Thus, in contrast with other toxins that contain only one or two receptor-binding domains^{16,17}, TcA contains in total four putative receptor-binding domains.

Remarkably, the amino- and carboxy-terminal regions of the two insertions that form the receptor-binding domains are located less than 8 Å apart (distance between 273 and 433, and between 1089 and 1764), suggesting that the inserted domains were added to an originally simpler shell. Indeed, we found that several homologous toxins from *P. luminescens* and other species lack some of the domains (Extended Data Table 1b). This might also explain the different host specificity of Tc toxins.

The pore-forming entity of TcA consists of a funnel formed by the TcB-binding domain (2328–2516) and a 200-Å-long α -helical central channel (Fig. 2a). Like the receptor-binding domains, the TcB-binding domain has an Ig-like β -sandwich fold (Extended Data Fig. 1). Each protomer contributes an intertwined helical pair that bends in a left-handed

¹Department of Physical Biochemistry, Max Planck Institute of Molecular Physiology, 44227 Dortmund, Germany. ²Institut für Experimentelle und Klinische Pharmakologie und Toxikologie, Albert-Ludwigs-Universität Freiburg, 79104 Freiburg, Germany. ³Department of Mechanistic Cell Biology, Max Planck Institute of Molecular Physiology, 44227 Dortmund, Germany. ⁴BIOSS Centre for Biological Signalling Studies, Albert-Ludwigs-Universität Freiburg, 79104 Freiburg, Germany. ⁵Institute of Chemistry and Biochemistry, Freie Universität Berlin, Thielallee 63, 14195 Berlin, Germany.

*These authors contributed equally to this work.

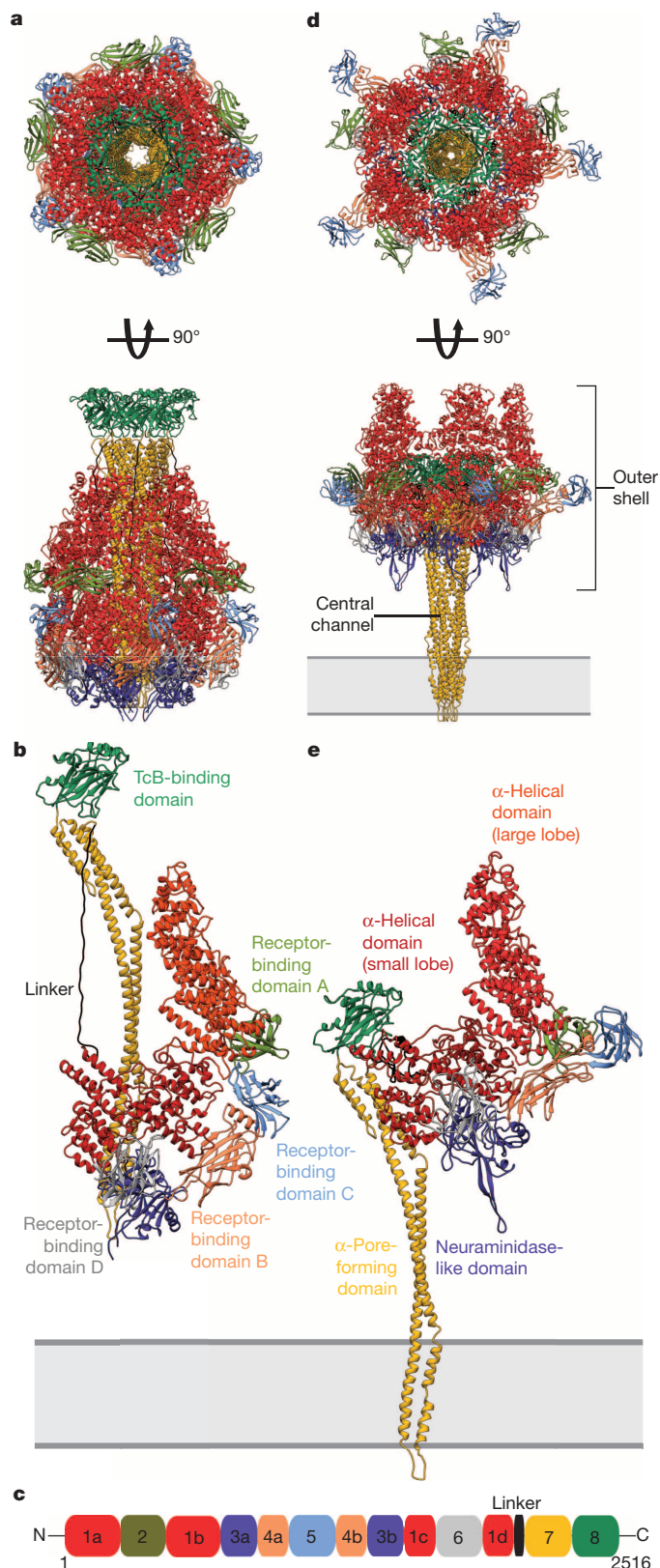


Figure 1 | Structures of the TcA prepore and pore complex. **a**, Top and side view of the TcA prepore pentamer. **b**, **c**, Side view of a TcA prepore protomer (**b**) and its domain organization (**c**). **d**, Top and side view of the TcA pore pentamer. **e**, Side view of a TcA pore protomer.

helical 180° turn over the length of the channel, forming an antiparallel coiled coil in the upper and middle part of the pore. At the top, the channel is further fortified by five helices (one per monomer) that

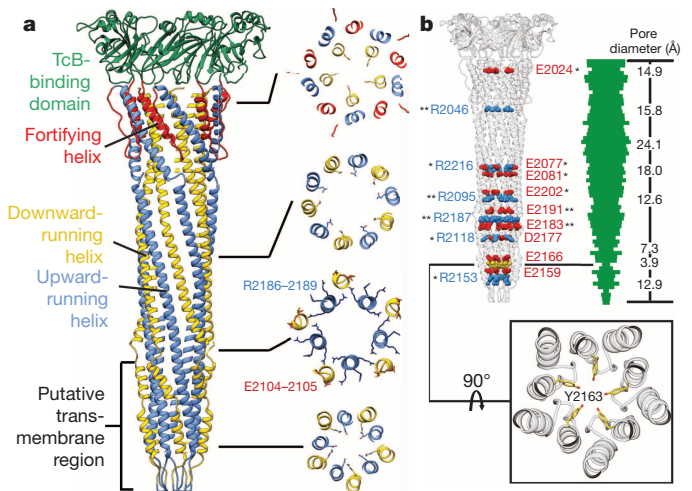


Figure 2 | The TcA prepore channel. **a**, The central channel is composed of ten long antiparallel α -helices that are interrupted by short unwound regions or loops. Helices running downwards and upwards are coloured yellow and blue, respectively. Helices that connect the channel with the funnel are depicted in red. **b**, Depiction of positively (blue), negatively (red) charged and aromatic (yellow) residues facing the interior of the channel. One and two asterisks depict highly conserved and conserved residues, respectively. The bar diagram (green) represents the inner diameter calculated using 3 Å steps. The numbers indicate the channel diameters at the respective height. The narrowest position of the channel is at tyrosine 2163 (yellow), which minimizes the diameter of the channel to 3.9 Å (see inset).

stabilize this region of the protein (Fig. 2a). At the bottom, the channel is closed by conserved β -turn loops formed by almost exclusively hydrophobic residues (2141–2153) (Fig. 2a and Extended Data Fig. 2a, b).

The outer surface of the lower end of the pore (pore surface residues between 2107–2158) is predominantly hydrophobic, confirming our previous prediction that this is the membrane-spanning part of the channel⁵ (Fig. 2a and Extended Data Fig. 2b). Notably, a ring of 15 arginines marks the upper end of the transmembrane region, indicating that this positively charged ring probably interacts with the negatively charged surface of the membrane and thereby stabilizes the pore within the lipid bilayer (Extended Data Fig. 2d).

Transition from the prepore to pore state

To reveal the mechanism of pore formation in molecular detail, we determined the structure of the TcA complex in its pore state by cryo-EM (Extended Data Fig. 3a–f) and built a molecular model derived from fitting the crystal structure of the TcA prepore into the cryo-EM structure (Fig. 1d, e and Extended Data Fig. 3g, h). Formation of the pore state is associated with marked conformational changes, resulting in the opening of the shell and a 12 nm shift of the central channel (Supplementary Video 2). Besides the linker between the shell and the channel, the conformations of the domains themselves are not considerably different from in the prepore state. However, their positions relative to each other change considerably, resulting in a completely different overall conformation (Fig. 1d, e). Comparing the prepore and pore structures, we identified three major hinge regions that are responsible for the opening of the shell (Extended Data Fig. 4a, b). Whereas two hinges involve receptor-binding domains, allowing their rearrangement (Extended Data Fig. 4c, d), a hinge between the two lobes of the α -helical shell domain is essential for the opening of the shell (Extended Data Fig. 4e). Notably, the arrangement of the four receptor-binding domains of TcA resembles the ectodomains of the interferon receptor IFNAR2 (ref. 18), suggesting that the different large-scale reorganization of these domains in the two TcA states represents an accommodation to different target geometries or an alteration of their binding affinity. The distance of receptor-binding domains to

the membrane is ~ 125 Å on average, suggesting that the receptor(s) on the cell surface probably have an elongated shape and protrude far from the membrane to bind to TcA, which is not unusual in eukaryotic cells. A good example of cell surface proteins that rise high above the membrane are integrins, which have an extracellular region of around 200 Å (ref. 19).

Entropic spring drives injection

The linker connecting the shell and the channel in the prepore state has a very unusual, stretched-out conformation, with a length of 113 Å for 48 residues (Fig. 3a). In the pore state, however, the linker appears to be contracted (Fig. 3b).

Steered molecular dynamics simulations of the transition between the two states indicate a gain in free energy of 20 – 66 kcal mol $^{-1}$, with a strong entropic contribution. This suggests that this process is exergonic and would therefore happen spontaneously, whether the collapsed state of the linker is folded or not (Fig. 3c). We therefore propose that the linker acts as an entropic spring. Comparable to a rubber band, it is stretched in the prepore state and can only achieve its preferred condensed conformation once the shell opens. This consequently pulls on the central channel and could therefore be the driving force for the previously described syringe-like injection mechanism of Tc toxins⁵ (Extended Data Fig. 4f and Supplementary Video 2).

The shell subunits from adjacent protomers overlap like an iris diaphragm, forming large interfaces in both TcA states. The interaction is mediated by electrostatic interactions involving different sets of residues in the prepore and pore state, respectively (Extended Data Fig. 2e, f). This indicates that both the prepore and pore state are stable conformations of the toxin complex. Interestingly, a shift to higher or lower pH values results in strong repulsions at the neuraminidase-like domains that close the shell at the bottom (Fig. 3d). We therefore propose that this domain acts as an electrostatic lock mainly responsible for the pH-induced opening of the shell⁵.

TcA translocation channel

The luminal surface of the channel is mainly negatively charged (Extended Data Fig. 2c), with many conserved charged residues protruding into the lumen of the channel (Fig. 2b), supporting our previous findings that the channel is cation selective²⁰. The cation selectivity of the anthrax translocation pore²¹ was shown to be a prerequisite for the unidirectional translocation of anthrax lethal factor (LF), making use of the pH gradient across the endosomal membrane²². It is thought that protonation before and deprotonation after the translocation of anionic proteins is the basis of a Brownian ratchet movement that drives the

translocation process in the direction of the compartment with the higher pH value, that is, the cytosol²³.

In contrast with LF, which is an anionic substrate, the translocated domains in Tc complexes from *P. luminescens*—that is, the hypervariable regions of TccC3 and TccC5—are both cationic substrates with isoelectric points of 9.68 and 8.65, respectively, and would therefore be translocated without additional pH gradients. This suggests a pH-independent mechanism of translocation in Tc toxins.

Interestingly, the diameter of the channel varies and opens at the height of a larger hydrophobic region, but generally narrows from the top to the bottom, where it is closed by the β -turn loops (Fig. 2b and Extended Data Fig. 2b). The narrowest diameter of the part of the channel that does not permeate the membrane is 12.6 Å, indicating that α -helices but not folded proteins would fit through. The putative transmembrane region, however, is narrower and almost closed at tyrosine 2163 (minimal diameter of 3.9 Å), indicating that membrane permeation of the toxin must induce conformational changes at this site, thus triggering the opening of the channel (Fig. 2b).

TcB–TcC forms a large cocoon

To understand better how TcB and TcC interact and to understand the detailed structural organization of these subunits, we solved the crystal structure of the TcB–TcC (TcdB2–TccC3 from *P. luminescens*) fusion protein at a resolution of 2.35 Å. TcB and TcC are built by large β -sheets that are wound in a continuous anticlockwise spiral, forming a large hollow cocoon which is closed by a distorted six-bladed non-symmetrical β -propeller that sits at a $\sim 45^\circ$ angle to the longitudinal axis of the cocoon and the N-terminal region of TcB (5–100) (Fig. 4a, b, Extended Data Figs 5, 6a–e and Supplementary Video 1). The structures are almost identical to the previously described structure of the YenB–YenC2 complex from the insect pathogen *Yersinia entomophaga*⁹, although the latter shares only 56.5% protein sequence identity with TcB and TcC (Supplementary Discussion and Extended Data Fig. 5e, f). We performed mutational studies and tested the proteolytic activity of TcC mutants *in vitro* (Extended Data Fig. 7a) and their toxicity by treating cells *in vivo* (Extended Data Fig. 7b), and found that TcC acts as an aspartyl autoprotease, similarly to YenC2, with aspartates 651 and 674 as the typical catalytic dyad (Extended Data Fig. 5a–d).

The inner surface of the TcB–TcC cocoon is mainly positively charged and contains large hydrophobic patches, which creates a hostile milieu for protein folding (Extended Data Fig. 5g–k). This indicates that the TcB–TcC cocoon provides an ideal microenvironment to harbour an unfolded protein and to shield it from the surroundings. It might even be able to directly unfold native proteins. In line with this, the cleaved

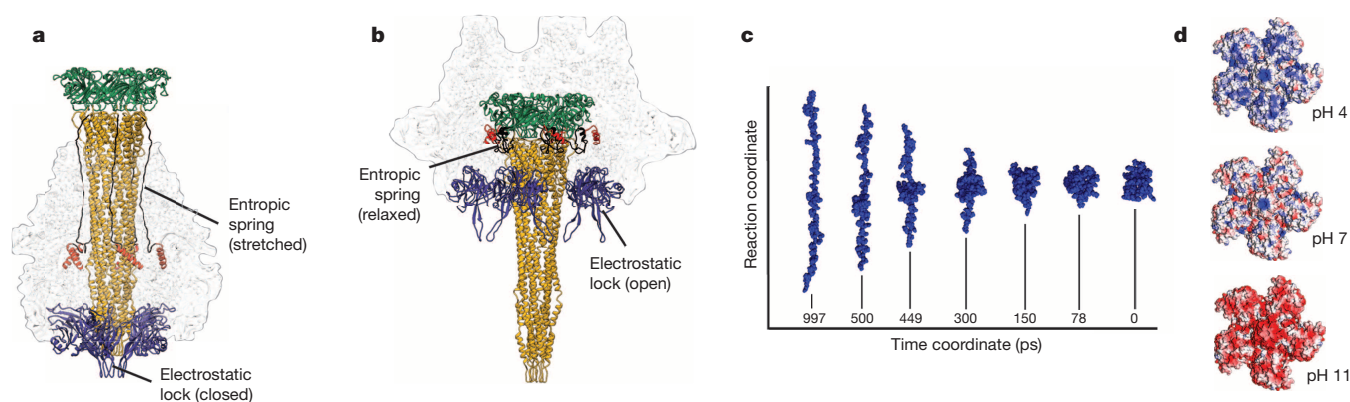
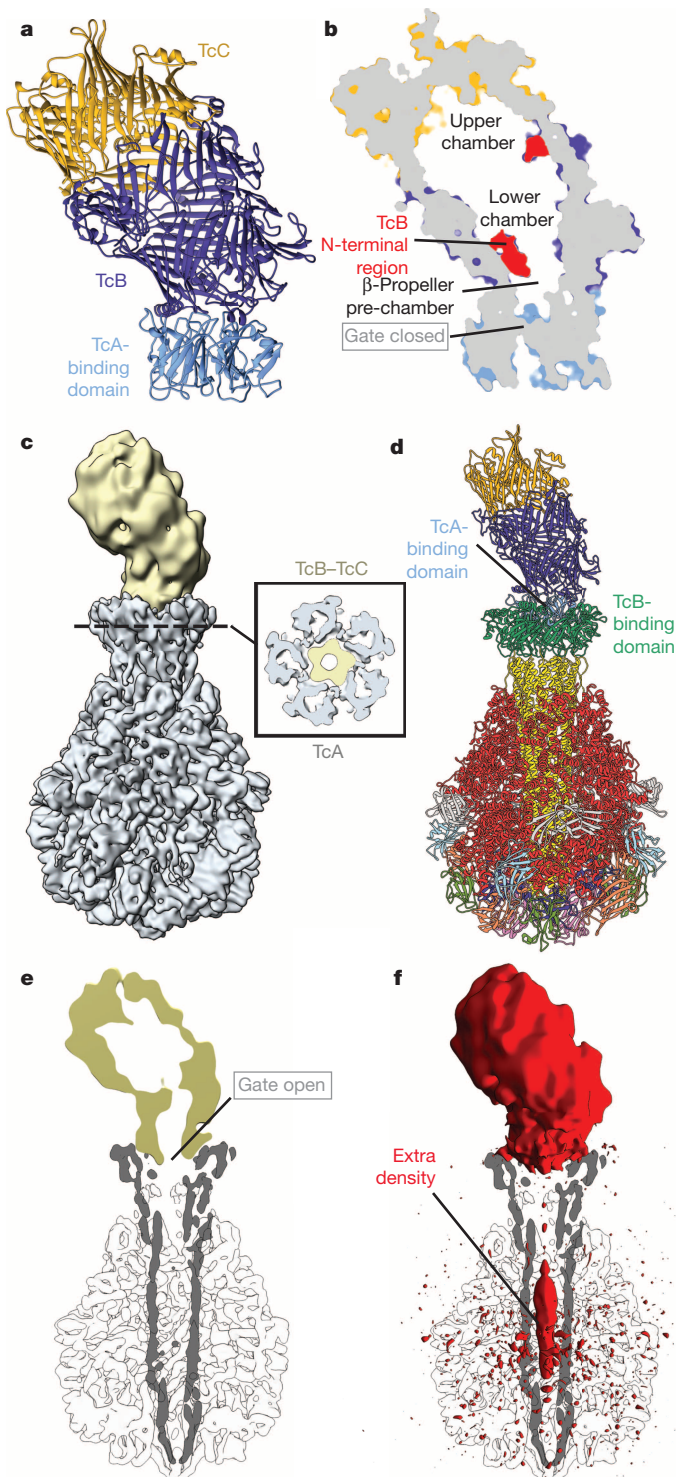


Figure 3 | An entropic spring drives the syringe-like injection mechanism of TcA. **a, b**, Structure of the TcA prepore (**a**) and pore complex (**b**), depicting the linker between the channel and shell domain (black) and the neuraminidase-like domain (blue). **c**, Steered molecular dynamics simulation, describing the contraction of the stretched linker, indicates that the linker is an entropic

spring. **d**, Surface electrostatic Coulomb potential of the neuraminidase-like domain at pH 4, 7 and 11. Strong repulsions at pH values of 4 and 11 but not at pH 7 indicate that the neuraminidase-like domain acts as an electrostatic lock that opens at high or low pH values.



off ADP-ribosyltransferase domain (679–960), which resides inside the cocoon before translocation (Extended Data Fig. 7c, d), is not resolved in our structure, indicating that the protein is either unfolded or in static disorder.

The TcA-binding domain acts as a gate

We then assembled the full holotoxin PTC3 prepore complex (TcdA1, TcdB2–TccC3) *in vitro* and determined its structure by cryo-EM (Fig. 4c and Extended Data Fig. 3i–l). The molecular model of PTC3 derived from fitting our crystal structures of the TcB–TcC and TcA complexes into the cryo-EM structure (Fig. 4d and Supplementary Videos 1, 3) clearly defines the interface between the two complexes

Figure 4 | Binding of TcB–TcC to TcA opens the β -propeller gate and the ADP-ribosyltransferase is secreted into the channel of TcA. **a**, TcB and TcC, depicted in blue and yellow, respectively, form a large cocoon. TcB also comprises the TcA-binding domain (light blue) that folds into a non-symmetrical six-bladed β -propeller. **b**, Central slice through TcB–TcC to visualize the upper and lower chamber as well as the β -propeller pre-chamber of the cocoon interior. The TcA-binding domain, TcB, TcC and the TcB N-terminal region are coloured in light blue, dark blue, yellow and red, respectively. **c**, Cryo-EM structure of the PTC3 (TcA/TcB–TcC) holotoxin complex. TcA and TcB–TcC are shown in grey and yellow, respectively. The inset shows a slice through the density at the indicated height of the interface between TcA and TcB–TcC, indicating that the β -propeller gate is open. **d**, Molecular model of the holotoxin complex, obtained by fitting the TcA crystal structure (Fig. 1a) and the TcB–TcC crystal structure (**a**) into the cryo-EM structure of the holotoxin complex shown in **c**. TcA domains are depicted in the same colours as in Fig. 1. TcB and TcC are coloured blue and yellow, respectively. The TcA-binding domain is shown in cyan. **e**, Slice through the density of the channel and TcB–TcC showing the open β -propeller gate and a continuous channel between the TcB–TcC cocoon and the channel of TcA. **f**, Difference map between the present cryo-EM structure of holotoxin and our previous cryo-EM structure of TcA alone (Electron Microscopy Data Bank accession EMD-2297). The difference map is shown in red overlaid with the central slice through the channel of the holotoxin shown in **e** depicted in grey.

(Extended Data Fig. 6f). The β -propeller of TcB and the domain of TcA that forms the highly conserved funnel of the central channel interact strongly with each other, identifying them as TcA- and TcB-binding domains, respectively (Fig. 4d).

Interestingly, the gate formed by the distorted β -propeller of TcB (Fig. 4b) is open in the holotoxin complex, resulting in a long continuous channel connecting the interior of the cocoon with that of the channel (Fig. 4e).

The density corresponding to the six-bladed β -propeller of TcB has a clear pseudo-five-fold symmetry (Fig. 4c and Extended Data Fig. 6g). This indicates that the conformational changes do not restore a six-fold symmetrical β -propeller, but that the β -propeller adapts to the symmetry mismatch by acquiring a pseudo-five-fold symmetry.

In the absence of complementary charges or matching hydrophobic patches at the interface (Extended Data Fig. 6f), shape complementarity and an induced fit are the basis for the specific interaction between TcB and TcA (Supplementary Video 3).

Most importantly, we find density inside the channel of TcA after holotoxin complex formation, indicating that the ADP-ribosyltransferase has passed through the β -propeller gate and entered the translocation pore before membrane permeation (Fig. 4f). The intensity of the density inside the pore is lower than the rest of the cryo-EM structure and fragmented, indicating that the protein is disordered inside the channel.

Thus, the binding of the distorted β -propeller of TcB to the pentameric funnel of TcA induces conformational changes that result in the rearrangement of the blades (Extended Data Fig. 6h and Supplementary Video 3) and thereby in the opening of the gate, allowing the passage of the toxic component into the target cell.

The major part of the open β -propeller pore through which the ADP-ribosyltransferase passes before it enters the TcA translocation channel is hydrophobic (Extended Data Fig. 6b). This region acts probably in a similar fashion to the Φ -clamp in anthrax toxin²⁴ and protects hydrophobic patches in the translocated protein. The narrow passage of the β -propeller also suggests that the ADP-ribosyltransferase must be unfolded before it passes through its gate, supporting our hypothesis that it is in an unfolded state inside the TcB–TcC cocoon.

The structures of TcA and TcB–TcC and of the holotoxin PTC3 described here reveal that Tc toxins are exceptional and function differently from any known toxins. Our data allow us to describe their mechanism of action in an unprecedented level of molecular detail (Supplementary Discussion and Extended Data Fig. 8). They form distinguished compartments for protein unfolding and processing, and translocate their toxic cargo in a novel way. Hence, our data lay a strong

foundation for understanding the function of other pathogens, including human pathogenic bacteria such as *Y. pestis* and *Y. pseudotuberculosis*. Furthermore, our data serve as a strong foundation for the development of protein shuttles targeting specific cell types in medical therapy.

METHODS SUMMARY

TcdA1 and TcdB2–TccC3 from *P. luminescens* were purified as described previously⁵. Crystals of TcA and TcB–TcC grew either in buffer containing 0.1 M HEPES pH 7.0, Jeffamine ED-2001 and 1.1 M sodium malonate or in buffer containing 0.1 M trisodium citrate pH 5.5, 0.1 M MgCl₂, 0.1 M NaCl and 12% PEG 4,000, respectively. X-ray diffraction data were collected at the PXII-X10SA beamline at the Swiss Light Source, and processed using the CCP4 program suite²⁵. The structure of TcA was solved at 4 Å using the previous 6.3 Å cryo-EM structure of TcA⁵ as a search model using PHENIX²⁶. The 2.35 Å structure of TcB–TcC was solved by experimental phasing. Images of vitrified TcA pore complexes and of the holotoxin (TcA, TcB–TcC) were taken with a JEOL JEM 3200FSC electron microscope equipped with a field-emission gun at an operation voltage of 200 kV. An in-column omega energy filter was used with a slit width of 15 eV. Micrographs were recorded with an 8k × 8k TemCam-F816 CMOS camera (TVIPS) under minimal dose conditions (15–20 e[−] per Å²). SPARX software²⁷ was used for all image-processing steps. For visualization and analysis we used Chimera and implemented programs²⁸. Steered molecular dynamics simulations were calculated using GROMACS²⁹.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 10 October 2013; accepted 10 January 2014.

Published online 23 February; corrected online 2 April 2014 (see full-text HTML version for details).

- Waterfield, N. R., Bowen, D. J., Fetherston, J. D., Perry, R. D. & French-Constant, R. H. The *tc* genes of *Photobacterium*: a growing family. *Trends Microbiol.* **9**, 185–191 (2001).
- French-Constant, R. H. & Bowen, D. J. Novel insecticidal toxins from nematode-symbiotic bacteria. *Cell. Mol. Life Sci.* **57**, 828–833 (2000).
- Hares, M. C. *et al.* The *Yersinia pseudotuberculosis* and *Yersinia pestis* toxin complex is active against cultured mammalian cells. *Microbiology* **154**, 3503–3517 (2008).
- Sheets, J. J. *et al.* Insecticidal toxin complex proteins from *Xenorhabdus nematophilus*: structure and pore formation. *J. Biol. Chem.* **286**, 22742–22749 (2011).
- Gatsogiannis, C. *et al.* A syringe-like injection mechanism in *Photobacterium luminescens* toxins. *Nature* **495**, 520–523 (2013).
- Bravo, A. & Sobrón, M. How to cope with insect resistance to Bt toxins? *Trends Biotechnol.* **26**, 573–579 (2008).
- French-Constant, R. H., Eleftherianos, I. & Reynolds, S. E. A nematode symbiont sheds light on invertebrate immunity. *Trends Parasitol.* **23**, 514–517 (2007).
- Lang, A. E. *et al.* *Photobacterium luminescens* toxins ADP-ribosylate actin and RhoA to force actin clustering. *Science* **327**, 1139–1142 (2010).
- Busby, J. N., Panjikar, S., Landsberg, M. J., Hurst, M. R. H. & Lott, J. S. The BC component of ABC toxins is an RHS-repeat-containing protein encapsulation device. *Nature* **501**, 547–550 (2013).
- Landsberg, M. J. *et al.* 3D structure of the *Yersinia entomophaga* toxin complex and implications for insecticidal activity. *Proc. Natl Acad. Sci. USA* **108**, 20544–20549 (2011).
- Lee, S. C. *et al.* Structural characterisation of the insecticidal toxin XptA1, reveals a 1.15 MDa tetramer with a cage-like structure. *J. Mol. Biol.* **366**, 1558–1568 (2007).
- Lesieur, C., Vécsey-Semjén, B., Abrami, L., Fivaz, M. & Gisou van der Goot, F. Membrane insertion: the strategies of toxins. *Mol. Membr. Biol.* **14**, 45–64 (1997).
- Murphy, J. R. Mechanism of diphtheria toxin catalytic domain delivery to the eukaryotic cell cytosol and the cellular factors that directly participate in the process. *Toxins* **3**, 294–308 (2011).
- Young, J. A. T. & Collier, R. J. Anthrax toxin: receptor binding, internalization, pore formation, and translocation. *Annu. Rev. Biochem.* **76**, 243–265 (2007).
- Varghese, J. N., Laver, W. G. & Colman, P. M. Structure of the influenza virus glycoprotein antigen neuraminidase at 2.9 Å resolution. *Nature* **303**, 35–40 (1983).
- Choe, S. *et al.* The crystal structure of diphtheria toxin. *Nature* **357**, 216–222 (1992).
- Petosa, C., Collier, R. J., Klimpel, K. R., Leppla, S. H. & Liddington, R. C. Crystal structure of the anthrax toxin protective antigen. *Nature* **385**, 833–838 (1997).
- Li, Z., Strunk, J. J., Lamken, P., Piehler, J. & Walz, T. The EM structure of a type I interferon–receptor complex reveals a novel mechanism for cytokine signaling. *J. Mol. Biol.* **377**, 715–724 (2008).
- Humphries, M. J. Integrin structure. *Biochem. Soc. Trans.* **28**, 311–339 (2000).
- Lang, A. E., Konukiewicz, J., Aktories, K. & Benz, R. TcdA1 of *Photobacterium luminescens*: electrophysiological analysis of pore formation and effector binding. *Biophys. J.* **105**, 376–384 (2013).
- Blaustein, R. O., Koehler, T. M., Collier, R. J. & Finkelstein, A. Anthrax toxin: channel-forming activity of protective antigen in planar phospholipid bilayers. *Proc. Natl Acad. Sci. USA* **86**, 2209–2213 (1989).
- Brown, M. J., Thoren, K. L. & Krantz, B. A. Charge requirements for proton gradient-driven translocation of anthrax toxin. *J. Biol. Chem.* **286**, 23189–23199 (2011).
- Feld, G. K., Brown, M. J. & Krantz, B. A. Ratcheting up protein translocation with anthrax toxin. *Protein Sci.* **21**, 606–624 (2012).
- Krantz, B. A. *et al.* A phenylalanine clamp catalyzes protein translocation through the anthrax toxin pore. *Science* **309**, 777–781 (2005).
- Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
- Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
- Hohn, M. *et al.* SPARX, a new environment for Cryo-EM image processing. *J. Struct. Biol.* **157**, 47–55 (2007).
- Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
- Berendsen, H. & van der Spoel, D. GROMACS: a message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **91**, 43–56 (1995).

Supplementary Information is available in the online version of the paper.

Acknowledgements We are grateful to R. S. Goody for continuous support and for useful comments on the manuscript. We thank V. Pfaumann for cloning, C. Schwan and S. Bergbrede for technical support and the Raunser laboratory for selecting particles. We thank the staff of beamline X10SA at the Paul Scherrer Institute, the X-ray communities at the Max Planck Institute (MPI) Dortmund and the MPI Heidelberg for help with data collection. This work was supported by the Deutsche Forschungsgemeinschaft, grants RA 1781/1-1 (to S.R.) and AK 6/22-1 (to K.A.), and by the Max Planck Society (to S.R.).

Author Contributions D.M. crystallized both protein complexes. D.M., R.G.E. and I.R.V. processed, refined and analysed X-ray data. C.G. and O.H. collected cryo-EM data; C.G. processed, refined and analysed cryo-EM data; A.E.L. purified proteins and performed mutational studies; I.R.V. calculated molecular dynamics simulations; K.A. and S.R. designed the study. D.M., C.G., R.G.E. and A.E.L. prepared figures. S.R. wrote the manuscript. All authors discussed the results and commented on the manuscript.

Author Information The coordinates for the EM structures have been deposited in the Electron Microscopy Data Bank under accession numbers EMD-2551 and EMD-2552. Coordinates of TcdB2–TccC3 and TcdA1 have been deposited in the Protein Data Bank under accession numbers 409X and 409Y, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.R. (stefan.raunser@mpi-dortmund.mpg.de).

Poly(A)-tail profiling reveals an embryonic switch in translational control

Alexander O. Subtelny^{1,2,3,4*}, Stephen W. Eichhorn^{1,2,3*}, Grace R. Chen^{1,2,3}, Hazel Sive^{2,3} & David P. Bartel^{1,2,3}

Poly(A) tails enhance the stability and translation of most eukaryotic messenger RNAs, but difficulties in globally measuring poly(A)-tail lengths have impeded greater understanding of poly(A)-tail function. Here we describe poly(A)-tail length profiling by sequencing (PAL-seq) and apply it to measure tail lengths of millions of individual RNAs isolated from yeasts, cell lines, *Arabidopsis thaliana* leaves, mouse liver, and zebrafish and frog embryos. Poly(A)-tail lengths were conserved between orthologous mRNAs, with mRNAs encoding ribosomal proteins and other ‘housekeeping’ proteins tending to have shorter tails. As expected, tail lengths were coupled to translational efficiencies in early zebrafish and frog embryos. However, this strong coupling diminished at gastrulation and was absent in non-embryonic samples, indicating a rapid developmental switch in the nature of translational control. This switch complements an earlier switch to zygotic transcriptional control and explains why the predominant effect of microRNA-mediated deadenylation concurrently shifts from translational repression to mRNA destabilization.

Most eukaryotic mRNAs end with poly(A) tails, which are added by a nuclear poly(A) polymerase following cleavage of the primary transcript during transcriptional termination¹. These tails are then shortened by deadenylases^{2,3}, although in some contexts (for example, animal oocytes or early embryos, or at neuronal synapses), they can be re-extended by cytoplasmic poly(A) polymerases^{4,5}. In the cytoplasm, the poly(A) tail promotes translation and inhibits decay^{2,5}.

Although poly(A) tails must exceed a minimal length to promote translation, an influence of tail length beyond this minimum is largely unknown. The prevailing view is that longer tails generally lead to increased translation^{5,6}. This idea partly stems from the known importance of cytoplasmic polyadenylation in activating certain genes in specific contexts^{4,5}, and the increased translation observed in *Xenopus* oocytes and *Drosophila* embryos when appending synthetic tails of increasing length onto an mRNA^{7,8}. Support for a more general coupling of tail length and translation comes from studies of yeast extracts⁹ and yeast cells^{10,11}. However, the general relationship between tail length and translational efficiency has not been reported outside of yeast, primarily because transcriptome-wide measurements have been unfeasible for longer-tailed mRNAs.

Poly(A)-tail length profiling by sequencing (PAL-seq)

We developed a high-throughput sequencing method that accurately measures individual poly(A) tails of any physiological length (Fig. 1a). After generating sequencing clusters and before sequencing, a primer hybridized immediately 3' of the poly(A) sequence is extended using a mixture of dTTP and biotin-conjugated dUTP as the only nucleoside triphosphates and conditions that were optimized to yield full-length extension products without terminal mismatches (Extended Data Fig. 1a). This key step quantitatively marks each cluster with biotin in proportion to the length of the poly(A) tail (Fig. 1a). After sequencing the 36 nucleotides immediately 5' of the poly(A) site, the flow cell is incubated with fluorophore-tagged streptavidin, which binds the biotin incorporated during primer extension to impart fluorescence intensity proportional to the poly(A)-tract length. To account for the density of each cluster, this raw intensity is normalized to that of the fluorescent

bases added during sequencing by synthesis¹², thereby yielding a normalized fluorescence intensity for the poly(A) tail of each transcript, paired with a sequencing read that identifies its poly(A) site and thus the gene of origin.

Each starting sample was spiked with a cocktail of mRNA-like standards of known tail lengths (Extended Data Fig. 1b) to produce a standard curve for converting normalized fluorescence intensities to poly(A)-tail lengths (Fig. 1b). We refer to each of these tail-length measurements paired with its identifying sequence as a poly(A) tag. Although recovery of tags from the standards varied somewhat, it did not vary systematically with tail length, which indicated that length-related biases were not an issue (Extended Data Fig. 1c). Additional analyses indicated that mRNA degradation did not bias against longer poly(A) tails (Extended Data Fig. 2a).

Because alternative start sites or alternative splicing can generate different transcripts with the same poly(A) site, we considered our results with respect to unique gene models (abbreviated as ‘genes’) rather than to transcripts (even though polyadenylation occurs on transcripts, not genes). Moreover, tags for alternative poly(A) sites of the same gene were pooled, unless stated otherwise. With this pipeline, analysis of RNA from NIH3T3 mouse fibroblasts (3T3 cells) yielded at least one tag from 10,094 unique protein-coding genes (including 97% of the 9,976 genes with at least one mRNA molecule per cell, as determined by RNA-seq) and ≥ 100 tags from 2,873 genes, coverage typical of most samples (Supplementary Table 1).

Tail-length diversity within each species

Median tail lengths in mammalian cells (range, 67–96 nucleotides) exceeded those in *A. thaliana* leaves and *Drosophila melanogaster* S2 cells (51 and 50 nucleotides, respectively), which exceeded those in budding (*Saccharomyces cerevisiae*) and fission (*Schizosaccharomyces pombe*) yeast (27 and 28 nucleotides, respectively) (Fig. 2a). Similar differences between mammalian, fly, plant and yeast cells were observed when comparing tail-length averages for individual genes (Fig. 2b). For genes within each species, mean tail lengths varied, with the 10th and 90th

¹Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ²Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA. ³Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ⁴Harvard-MIT Division of Health Sciences and Technology, Cambridge, Massachusetts 02139, USA.

*These authors contributed equally to this work.

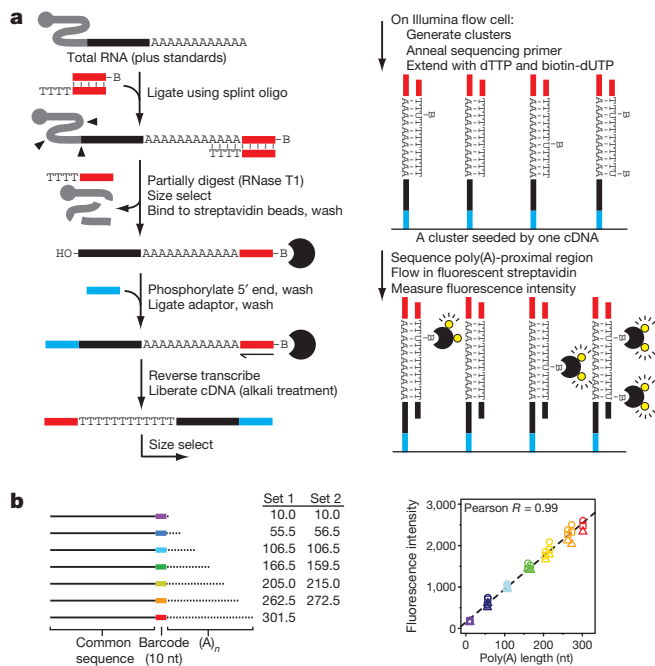


Figure 1 | Global measurement of poly(A)-tail lengths. **a**, Outline of PAL-seq. For each cluster, the fluorescence intensity reflects the tail length of the cDNA that seeded the cluster. Although the probability of incorporating a biotin-conjugated dU opposite each tail nucleotide is uniform, stochastic incorporation results in a variable number of biotins for each molecule within a cluster. **b**, Median streptavidin fluorescence intensities for two sets of mRNA-like molecules with indicated poly(A)-tail lengths, which were added to 3T3 (circle), HEK293T (triangle) and HeLa (square) samples for tail-length calibration.

percentiles differing by 1.4- to 1.6-fold. Variation was also observed for different mRNA transcripts from the same gene (Fig. 2c). For most genes the distributions were unimodal, with the mode approaching the mean (Fig. 2d). Poly(A)-tail lengths increased when progressing through

the cleavage, blastula and gastrula stages of zebrafish embryonic development (2, 4 and 6 h post-fertilization (hpf), respectively) and the analogous stages of frog development (Fig. 2a, b, d). Processed data reporting tail lengths for all genes detected in each sample are provided in the Gene Expression Omnibus (accession number GSE52809).

Comparison of tail lengths for orthologous genes in human (HeLa and HEK293T) and mouse (3T3 and liver) cells revealed moderately strong correlations, indicating that tail lengths are conserved (Extended Data Table 1, Spearman R_s as high as 0.46). When searching for gene classes that tended to have longer or shorter tails, the most striking and pervasive enrichment was for ribosomal protein and other 'housekeeping' genes among the short-tailed genes (Extended Data Table 2). This enrichment was strong in yeast, despite previous reports that ribosomal protein genes tend to have long tails^{10,11}. To address this and other discrepancies with previous yeast studies (Extended Data Fig. 3a, b), we used an independent method to measure the poly(A)-tail lengths of eight yeast genes, including four ribosomal protein genes. The results were much more consistent with our measurements than with the previous measurements (Extended Data Figs 3 and 4). Both previous reports used the polyadenylation state microarray (PASTA) method, which fractionates RNAs by stepwise thermal elution from poly(U)-Sephacrose. Although studies have successfully used poly(U)-Sephacrose fractionation to detect tail-length changes for the same genes in different contexts^{13–15}, detecting differences between different genes in the same context is more challenging. Our results suggest that PASTA, as previously implemented in yeasts^{10,11}, is less suitable than PAL-seq for intergenic comparisons, although we cannot exclude the possibility that the discrepancies arose from different growth conditions.

The types of genes with shorter or longer tails differed between the embryonic samples and the other samples (Extended Data Table 2). Genes in the early embryo might not have the same tail lengths as their orthologues do in other contexts because before the maternal-to-zygotic transition (MZT), which occurs at ~ 3 hpf in zebrafish¹⁶ and at approximately stage 8 in *Xenopus laevis*¹⁷, transcription is not yet active, and some maternal transcripts are masked for later use, whereas others are subject to cytoplasmic polyadenylation⁵. At 6 hpf in zebrafish, ribosomal protein mRNAs had switched from being enriched in shorter-tailed genes

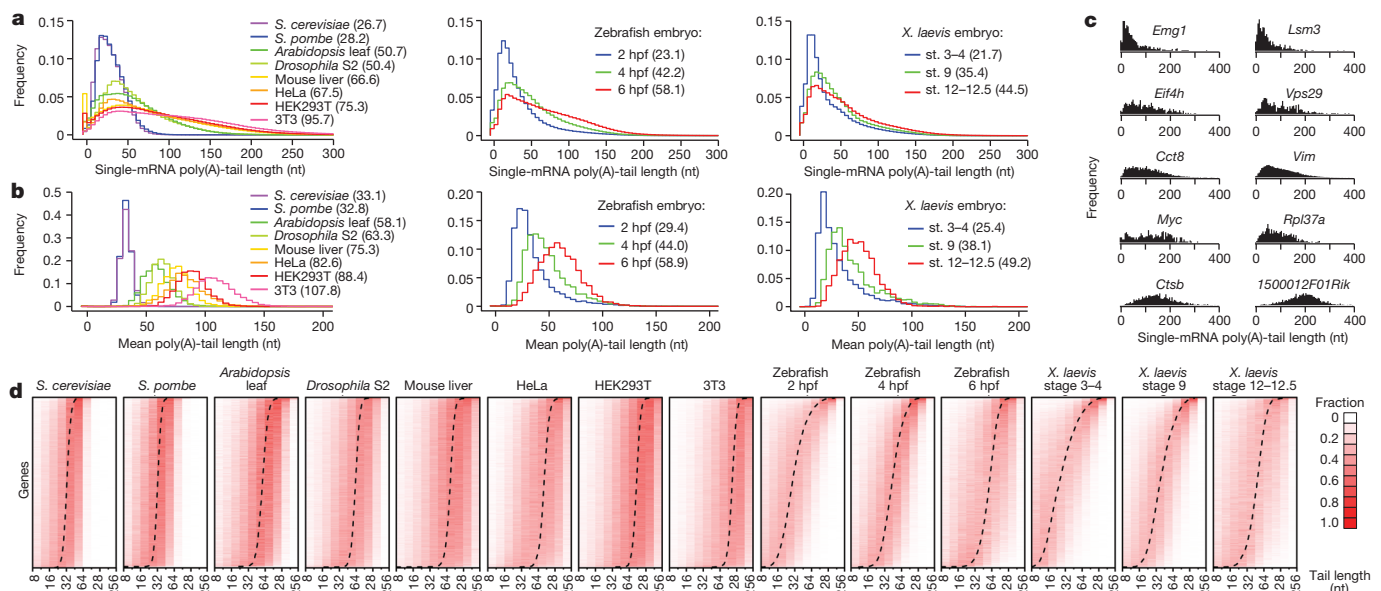


Figure 2 | Poly(A)-tail lengths in yeast, plant, fly and vertebrate cells. **a**, Bulk tail-length distributions. For each sample, histograms tally tail-length measurements for all poly(A) tags mapping to annotated 3' UTRs (bin size = 5 nucleotides). Leftmost bin includes all measurements <0 nucleotides. Median tail lengths are in parentheses. **b**, Intergenic tail-length distributions. For each sample, histograms tally average tail lengths for protein-coding genes with ≥ 5 tags (yeasts, zebrafish and *Xenopus*) or ≥ 100 tags (other samples).

Median average tail lengths are in parentheses. **c**, Intragenic tail-length distributions for 10 genes sampling the spectrum of average tail lengths in 3T3 cells. **d**, Intragenic tail-length distributions. Heat maps show the frequency distribution of tail lengths for each gene tallied in **b**. The colour intensity indicates the fraction of the total for the gene. Genes are ordered by average tail length (dashed line). Results from the *S. cerevisiae* total-RNA sample are reported in this figure.

to being enriched in longer-tailed genes (Extended Data Table 2), perhaps because these were mostly newly synthesized transcripts, which tended to have longer tails at this stage (Extended Data Fig. 5).

Because deadenylation is an important early step in eukaryotic mRNA decay^{2,3,18}, we examined the relationship between poly(A)-tail length and published mRNA stability values (Extended Data Table 1). Tail length and half-life were slightly negatively correlated in HeLa and 3T3 cells ($R_s = -0.048$ and -0.16 , respectively) and variably correlated in yeast, depending on the source of the half-life measurements (R_s from -0.44 to 0.23). The weak relationships in HeLa and 3T3 cells would be expected if mRNAs with different half-lives have similar steady-state tail-length distributions, with the less stable mRNAs transiting through the distributions more quickly.

No strong, easily interpretable correlations between tail length and mRNA features (length of 3' untranslated region (3' UTR), length of open reading frame (ORF), total length, splice-site number, splice-site density) or expression (steady-state accumulation and nuclear-to-cytoplasmic ratio) were observed (Extended Data Table 1). Of these, the strongest correlations were between tail length and steady-state accumulation (R_s from -0.44 to 0.25), and between tail length and mRNA length (R_s from -0.12 to 0.36) or features related to mRNA length. Support for the latter relationship was also observed in intragenic comparisons, which revealed a weak positive relationship between tail length and the length of tandem 3'-UTR isoforms (Extended Data Fig. 6a). In early zebrafish embryos this relationship between 3'-UTR isoforms was even more pronounced when a predicted cytoplasmic polyadenylation element (CPE)^{4,19} was present in the region unique to the longer isoform (Extended Data Fig. 6b).

Decoupling of tail length and translation

Most reports of increased translation of longer-tailed mRNAs have used oocytes and early embryos^{4,5}. To examine whether this phenomenon reported in early embryos for a few genes applies transcriptome-wide, we performed ribosome footprint profiling and RNA-seq to measure translational efficiencies²⁰ from the embryonic samples used to measure tail lengths. We found that in early embryos (cleavage and blastula stages) of both fish and frog, mean poly(A)-tail length correlated strongly with translational efficiency (Fig. 3a, R_s from 0.62 to 0.77). No other mRNA feature has been reported to correlate so well with translational efficiency in any system.

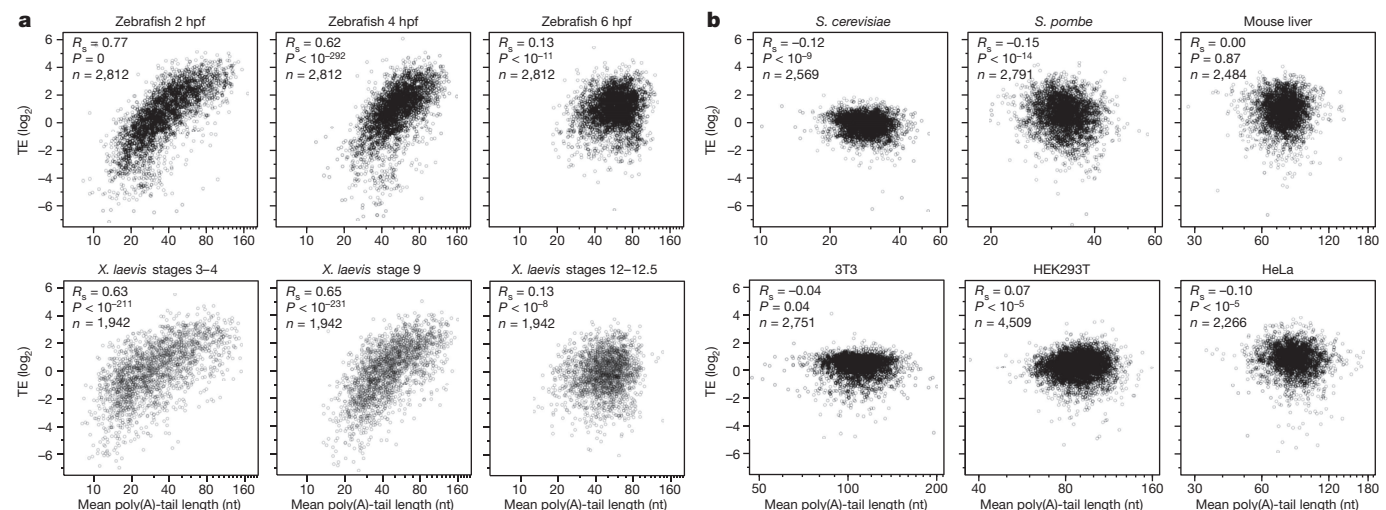


Figure 3 | Transient coupling between poly(A)-tail length and translational efficiency. **a**, Relationship between mean tail length and translational efficiency (TE) for genes with ≥ 50 poly(A) tags from embryonic samples at the indicated developmental stages. For each stage, tail lengths and translational efficiencies were obtained from the same sample. *MGC116473* and *DDX24* fell outside the plot for *X. laevis*, stages 3–4, and *LOC100049092* fell outside the plot for *X. laevis*, stages 12–12.5. **b**, Relationship between mean tail length and

translational efficiency in the indicated cells, for genes with ≥ 50 (yeasts) or ≥ 100 (others) tags. With the exception of HeLa³⁵, tail lengths and translational efficiencies were from the same samples. *S. cerevisiae* YBR196C, YLR355C and YDL080C, *S. pombe* SPCC63.04.1, mouse liver NM_007881 and NM_145470, HEK293T NM_001007026, NM_021058 and NM_003537, and HeLa NM_001007026 fell outside their respective plots.

In these early embryonic stages, a twofold increase in tail length corresponded to a large increase in translational efficiency—greater than 6-fold when doubling the tail from 20 to 40 nucleotides in 2 hpf zebrafish (Fig. 3a). Although longer-tailed mRNAs were more likely to contain a CPE, the relationship between tail length and translational efficiency for CPE-containing mRNAs was no different from that of other mRNAs (Extended Data Fig. 7a). In theory, this coupling might not be causal, or it might be causal but strictly due to either translational inhibition causing tail shortening or translational activity preventing tail shortening. Alternatively, all or at least some of the coupling might result from longer tail length causing more efficient translation in the early embryo. We favour this last possibility because it agrees with the known importance of cytoplasmic polyadenylation for activating genes in maturing oocytes^{8,21,22} and early embryos^{23,24} of *Xenopus* and in certain other vertebrate contexts^{25–30}. Even more importantly, it agrees with the increased translation observed in *Xenopus* oocytes when appending prosthetic poly(A) tails of increasing length onto an mRNA⁸.

Intragenic comparison of tail length and translation

The simplest interpretation of the weak or negative correlations we observed between tail length and translational efficiency in yeast and mammalian cells is that increasing average tail length over the physiological range does not enhance translation in these contexts. However, our comparisons of average tail length and average translational efficiency

between genes (Fig. 3b) might have missed a relationship that would be observed when looking at differentially translated mRNAs from the same gene. To address this possibility, we fractionated 3T3 cell lysate to isolate mRNAs associated with different numbers of ribosomes and measured the tail lengths in each fraction (Fig. 4a). To learn how poly(A)-tail length related to ribosome density for individual genes, we plotted mean tail-length values as a function of the number of bound ribosomes and fit the data for each gene with a straight line (Fig. 4b). The slopes of these lines were generally small, and most were slightly negative (Fig. 4b); positive slopes would have been expected if longer tails enhanced translation. Thus, the increase in median length observed between the lightest and heaviest fractions when considering bulk tail lengths (Fig. 4a; 66 and 82 nucleotides, respectively) did not indicate a relationship between longer tails and enhanced translation but instead might have reflected the positive correlation between ORF length and tail length observed in 3T3 cells (Extended Data Table 1; $R_s = 0.36$). The trend of mostly negative slopes prevailed even when excluding data from mRNA not associated with any ribosomes (Extended Data Fig. 7c), or when examining subsets of genes with longer or shorter mean tail lengths, or with higher or lower translational efficiencies (Extended Data Fig. 7d). This global intragenic analysis (Fig. 4b) supports the conclusion drawn from intergenic analyses (Fig. 3), that in all yeast and mammalian contexts examined (and presumably in most other cellular contexts), mRNAs with longer poly(A) tails are not more efficiently translated.

A shift in the ultimate effects of miRNAs

MicroRNAs (miRNAs) are small RNAs that pair to sites in mRNAs to target these messages for post-transcriptional repression³². Global measurements indicate that miRNA targeting causes mostly mRNA destabilization, with translational repression comprising a detectable but minor component of the overall repression^{33–36}. The only known exception is the transient translational repression observed in early zebrafish embryos³⁶. At 4 hpf miR-430 targeting causes mostly translational repression with very little mRNA destabilization, whereas by 6 hpf the outcome shifts to mostly mRNA destabilization³⁶. Because miR-430 is induced only ~1.5 h before the 4-hpf stage, these results are interpreted as revealing the dynamics of miRNA action, in which an early phase of translational repression gives way to a later phase in which destabilization dominates³⁶. When considering that miRNA

targeting promotes poly(A)-tail shortening through the recruitment of deadenylase complexes³⁷, our results suggest an alternative mechanism for the shift in miRNA regulatory outcomes. In this mechanism, miRNAs mediate tail shortening at both 4 and 6 hpf, but because of the switch in the nature of translational control (as well as destabilization of short-tailed mRNAs at later stages), tail shortening has very different consequences in the two stages: at 4 hpf, tail shortening predominantly decreases translational efficiency, whereas at 6 hpf, it predominantly decreases mRNA stability.

To integrate miRNA-mediated repression with effects on tail length, we injected one-cell zebrafish embryos with miRNAs that are normally not present in the early embryo and examined the influence of these injected miRNAs on ribosome-protected fragments, mRNA levels and poly(A)-tail lengths at 2, 4 and 6 hpf. Injecting miR-155 caused ribosome-protected fragments from many of its predicted targets to decrease relative to ribosome-protected fragments from no-site control mRNAs (Fig. 5a). Despite the decrease in ribosome-protected fragments, target mRNA levels did not change relative to the controls at 2 and 4 hpf, indicating that at these stages miR-155 targeting caused mostly translational repression. In contrast, decreases in ribosome-protected fragments were accompanied by nearly commensurate mRNA reductions at 6 hpf, indicating that by this stage the outcome of repression had shifted to mostly mRNA destabilization (Fig. 5a). Thus, the shift in miRNA regulatory outcome that occurs between 4 and 6 hpf is not specific to miR-430 or its targets. With respect to mechanism, the observation of this shift between 4 and 6 hpf, even though the injected miR-155 was present and active much earlier than was miR-430, indicated that the shift reflected a transition from the unusual regulatory regime operating in pre-gastrulation embryos (in which translational efficiency is sensitive to tail length) more than it reflected the dynamics of miRNA action.

The tail-length results further supported a mechanism involving shifting consequences of tail-length shortening. Predicted miR-155 targets had shortened tails at 2 and 4 hpf (Fig. 5b), which explained most of the miRNA-induced translational repression observed at these stages (Fig. 5c). By 6 hpf, the tail-length decreases observed at 4 hpf had mostly abated for predicted miR-155 targets (Fig. 5b), and these mRNAs were instead less abundant (Fig. 5a), in concordance with their extent of deadenylation at 4 hpf (Extended Data Fig. 8a). These observations agreed with the idea that tail shortening destabilizes mRNAs at later developmental stages and indicated that the miRNA-mediated deadenylation occurring during the earlier developmental stages promotes decay later. With shorter tails no longer associated with reduced translation (Fig. 3a) and instead associated with reduced mRNA levels, the ultimate consequence of miRNA-mediated repression shifted from translational repression to mRNA destabilization (Fig. 5c). Analogous results were obtained after injecting a different miRNA, miR-132 (Fig. 5c, Extended Data Fig. 8).

Because tail length was no longer strongly coupled with translational efficiency (Fig. 3a), tail-length changes did not explain the decrease in mean translational efficiency observed at 6 hpf for predicted miR-132 targets (Fig. 5c). We conclude that when poly(A)-tail length is uncoupled from translational efficiency, the translational repression often detected as a minor component of the overall repression^{33–35} arises from a mechanism different from the one that dominates pre-gastrulation.

Our results provide a compelling explanation for miRNA-mediated translational repression in the pre-gastrulation zebrafish embryo: miRNAs induce poly(A) shortening, which decreases translational efficiency at this developmental period. They also explain why the pre-gastrulation zebrafish embryo is the only known context for which translational repression is the dominant outcome of miRNA-mediated regulation; in all other contexts examined, tail-length shortening causes mRNA destabilization with little or no effect on translational efficiency.

Two gene-regulatory regimes

Our results from yeast, cultured mammalian cells and mouse liver refute the prevailing view that poly(A)-tail length broadly influences

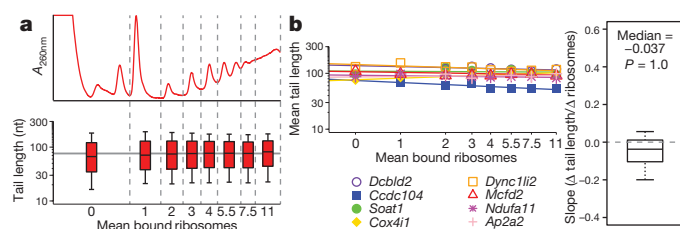


Figure 4 | No detectable intragenic coupling between poly(A)-tail length and translational efficiency. **a**, Global analysis of tail lengths across a polysome profile for 3T3 cells. The absorbance trace indicates mean number of ribosomes bound per mRNA for each fraction from the sucrose gradient (top, fractions demarcated with vertical dashed lines). Box plots show distributions of bulk tail lengths in each fraction for all tags mapping to annotated 3' UTRs (bottom). Box plot percentiles are line, median; box, 25th and 75th percentiles; whiskers, 10th and 90th percentiles. The horizontal line indicates the overall median of the median tail lengths. **b**, Relationship between tail lengths and ribosomes bound per mRNA for mRNAs from the same gene. For each gene, the data from **a** were used to plot the mean tail length as a function of bound ribosomes. Log-log plots for 8 randomly selected genes with ≥ 50 poly(A) tags in ≥ 6 fractions are shown (left), with lines indicating linear least-squares fits to the data (adding a pseudocount of 0.5 ribosomes to the fraction with 0 ribosomes). The box plot shows the distribution of slopes for all genes with ≥ 50 poly(A) tags in ≥ 4 fractions (right; $n = 4,079$; one-sided, one-sample Wilcoxon test; box plot percentiles as in **a**).

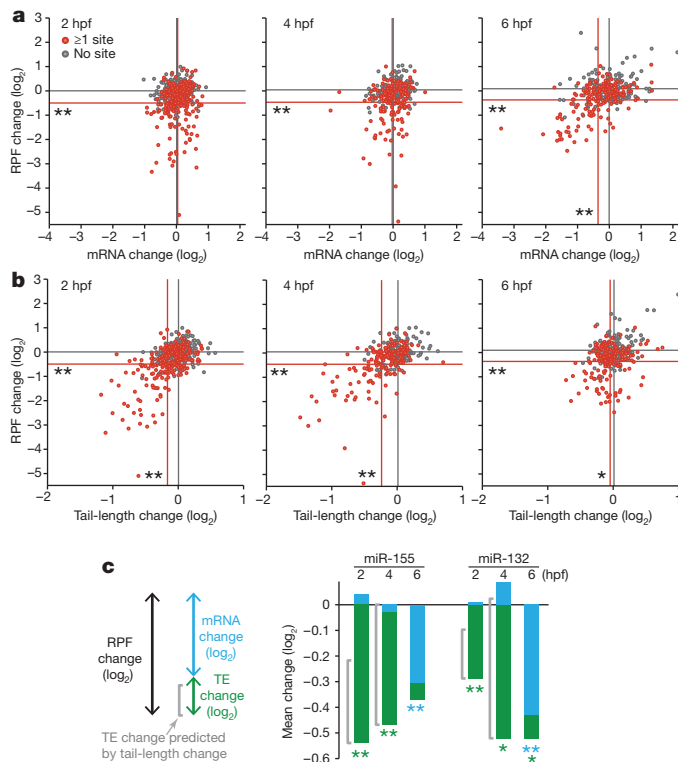


Figure 5 | The influence of miR-155 on ribosomes, mRNA abundance and tails in the early zebrafish embryo. **a**, Relationship between changes in ribosome-protected fragments (RPFs) and changes in mRNA levels after injecting miR-155. Changes observed between miRNA- and mock-injected embryos are plotted at the indicated stages for predicted miR-155 target genes (red, genes with ≥ 1 miR-155 site in their 3' UTR) and control genes (grey, genes that have no miR-155 site, yet resemble the predicted targets with respect to 3'-UTR length). To ensure that differences observed between 4 and 6 hpf were not the result of examining different genes, only site-containing genes and no-site control genes detected at both 4 and 6 hpf are shown for these stages. Lines indicate mean changes for the respective gene sets, with statistically significant differences between the sets indicated ($*P \leq 0.05$; $**P < 10^{-4}$, one-tailed Kolmogorov–Smirnov test). Because injected miRNAs partially inhibited miR-430-mediated repression, genes with miR-430 sites were not considered. Data were normalized to the median changes observed for the controls. **b**, Relationship between changes in ribosome-protected fragments and changes in mean tail lengths after injecting miR-155. Tail lengths were determined using PAL-seq, otherwise as in **a**. **c**, A developmental switch in the dominant mode of mRNA-mediated repression. The schematic (left) depicts the components of the bar graphs, showing how the changes in ribosome-protected fragments (RPFs) comprise both mRNA and translational efficiency (TE) changes. The compound bar graphs show the fraction of repression attributed to mRNA degradation (blue) and translational efficiency (green) for the indicated stage, depicting the overall impact of miR-155 (centre; plotting results from **a** and **b** for genes with sites) and miR-132 (right, plotting results from Extended Data Fig. 8b for genes with sites). Slight, statistically insignificant increases in mRNA for predicted targets resulted in blue bars extending above the axis. For samples from stages at which tail length and translational efficiency were strongly coupled, a bracket adjacent to the compound bar indicates the fraction of repression attributable to shortened tails. Significant changes for each component are indicated with asterisks of the corresponding colour ($*P \leq 0.05$; $**P < 10^{-4}$, one-tailed Kolmogorov–Smirnov test).

translational efficiency. In doing so, they add to the known differences between the regulatory regime operating in these cells and that operating in early metazoan embryos.

This absence or presence of coupling between poly(A)-tail length and translational efficiency can be rationalized in light of the potential interplay among regulatory options available in the two regulatory regimes. Our yeast, mammalian and gastrulation-stage cells were transcriptionally

active, which offers ample opportunities for nuclear control of gene expression. Moreover, active transcription enables unstable mRNAs to be replaced if required, thereby expanding the contexts in which differential mRNA stability can be exploited for gene control. Thus, an additional layer of control in which translational efficiency depends on poly(A)-tail length is dispensable. More importantly, because this type of coupling would lower output from older mRNA molecules that, in the absence of cytoplasmic polyadenylation, would often have shorter poly(A) tails, the utility of gene regulation through mRNA stability would be compromised. In this conventional regulatory regime, long-lived mRNAs would have less value if they were translated less efficiently because of their shorter tails.

For fish and frog embryos at the cleavage stage, the regulatory regime was very different. These embryos were transcriptionally inactive, which not only precludes the use of transcription and other nuclear processes to alter gene expression programs but also limits the use of differential mRNA stability, because degraded mRNAs cannot be replaced until zygotic transcription begins. Perhaps as a consequence, many mRNAs with short tails were observed (Fig. 2a), consistent with the known stability of short-tailed mRNAs in early embryos^{19,38}. In these circumstances, early embryonic cells apparently harness differential tail length for global gene control. This result expands the known behaviour of individual genes in *Xenopus* embryos^{23,24} and the observation that early embryonic cells have robust cytoplasmic polyadenylation⁴, which increases the utility of a tail-length regulatory mechanism. Compared to metazoan cells subject to the standard regulatory regime (for example, 6-hpf zebrafish embryos and the mammalian cells examined), cleavage-stage embryos had more uniform intragenic tail lengths and more variable intergenic lengths (Fig. 2d), as required for efficient harnessing of the tail-length regulatory regime. With their tail-length distribution also shifted towards shorter tails (Fig. 2b), cleavage-stage embryos can most efficiently exploit the tail-length differences with the greatest impact (Fig. 3a).

The transition between these two very different gene-regulatory regimes was rapid but not immediate. Despite their zygotic transcription, late-blastula embryos still coupled tail length with translation. Indeed, to the extent that newly transcribed zygotic mRNAs tended to have longer tails than did the maternally inherited mRNAs (Extended Data Fig. 5), the continued coupling observed in this hybrid regime would act to increase the relative output from these newly minted mRNAs, thereby sharpening the MZT.

We suspect that the tail-length regulatory regime observed in early embryos operates in other systems in which transcription is repressed (or occurs at a distant location) and cytoplasmic polyadenylation is active, such as early embryos of other metazoan species, maturing oocytes and neuronal synapses⁵. The ability to measure poly(A)-tail lengths at single-mRNA resolution should provide important insights in these systems.

METHODS SUMMARY

Cytoplasmically enriched lysates were prepared from HEK293T, 3T3, mouse liver, *X. laevis*, *S. pombe* and zebrafish samples, as well as one of the *S. cerevisiae* samples, and divided into three portions, one each for PAL-seq, RNA-seq and ribosome profiling. PAL-seq was performed as outlined in Fig. 1a. RNA-seq and ribosome profiling were performed essentially as described previously³⁵. Poly(A) tags were mapped to a reference genome (or transcriptome) of the species, carrying forward those that mapped uniquely to the genome (or transcriptome) and also overlapped with the 3' UTR of a transcript model chosen to represent a gene. Ribosome-protected fragments and RNA-seq tags were mapped to ORFs, as described previously³⁵, except tags mapping within the first 50 nucleotides of an ORF were discarded (to exclude signal from ribosomes that might have initiated after cycloheximide was added). Each mRNA with a 3' UTR that had at least one 7-nucleotide site matching the miRNA seed region³² was predicted to be a target of that miRNA. Genes that had no 6-nucleotide seed match anywhere within their mature transcript were classified as no-site genes, from which a set of no-site control genes was selected such that its 3'-UTR length distribution matched that of the predicted targets.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 25 July; accepted 23 December 2013.

Published online 29 January 2014.

1. Moore, M. J. & Proudfoot, N. J. Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell* **136**, 688–700 (2009).
2. Goldstrohm, A. C. & Wickens, M. Multifunctional deadenylase complexes diversify mRNA control. *Nature Rev. Mol. Cell Biol.* **9**, 337–344 (2008).
3. Chen, C. Y. & Shyu, A. B. Mechanisms of deadenylation-dependent decay. *Wiley Interdiscip. Rev. RNA* **2**, 167–183 (2011).
4. Richter, J. D. Cytoplasmic polyadenylation in development and beyond. *Microbiol. Mol. Biol. Rev.* **63**, 446–456 (1999).
5. Weill, L., Belloc, E., Bava, F. A. & Mendez, R. Translational control by changes in poly(A) tail length: recycling mRNAs. *Nature Struct. Mol. Biol.* **19**, 577–585 (2012).
6. Eckmann, C. R., Rammelt, C. & Wahle, E. Control of poly(A) tail length. *Wiley Interdiscip. Rev. RNA* **2**, 348–361 (2011).
7. Sallés, F. J., Lieberfarb, M. E., Wreden, C., Gergen, J. P. & Strickland, S. Coordinate initiation of *Drosophila* development by regulated polyadenylation of maternal messenger RNAs. *Science* **266**, 1996–1999 (1994).
8. Barkoff, A., Ballantyne, S. & Wickens, M. Meiotic maturation in *Xenopus* requires polyadenylation of multiple mRNAs. *EMBO J.* **17**, 3168–3175 (1998).
9. Preiss, T., Muckenthaler, M. & Hentze, M. W. Poly(A)-tail-promoted translation in yeast: implications for translational control. *RNA* **4**, 1321–1331 (1998).
10. Beilharz, T. H. & Preiss, T. Widespread use of poly(A) tail length control to accentuate expression of the yeast transcriptome. *RNA* **13**, 982–997 (2007).
11. Lackner, D. H. et al. A network of multiple regulatory layers shapes gene expression in fission yeast. *Mol. Cell* **26**, 145–155 (2007).
12. Nutiu, R. et al. Direct measurement of DNA affinity landscapes on a high-throughput sequencing instrument. *Nature Biotechnol.* **29**, 659–664 (2011).
13. Rosenthal, E. T., Tansey, T. R. & Ruderman, J. V. Sequence-specific adenylations and deadenylations accompany changes in the translation of maternal messenger RNA after fertilization of *Spisula* oocytes. *J. Mol. Biol.* **166**, 309–327 (1983).
14. Palatnik, C. M., Wilkins, C. & Jacobson, A. Translational control during early *Dictyostelium* development: possible involvement of poly(A) sequences. *Cell* **36**, 1017–1025 (1984).
15. Paynton, B. V., Rempel, R. & Bachvarova, R. Changes in state of adenylation and time course of degradation of maternal mRNAs during oocyte maturation and early embryonic development in the mouse. *Dev. Biol.* **129**, 304–314 (1988).
16. Kane, D. A. & Kimmel, C. B. The zebrafish midblastula transition. *Development* **119**, 447–456 (1993).
17. Newport, J. & Kirschner, M. A major developmental transition in early *Xenopus* embryos: II. Control of the onset of transcription. *Cell* **30**, 687–696 (1982).
18. Decker, C. J. & Parker, R. A turnover pathway for both stable and unstable mRNAs in yeast: evidence for a requirement for deadenylation. *Genes Dev.* **7**, 1632–1643 (1993).
19. Aanes, H. et al. Zebrafish mRNA sequencing deciphers novelties in transcriptome dynamics during maternal to zygotic transition. *Genome Res.* **21**, 1328–1338 (2011).
20. Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. & Weissman, J. S. Genome-wide analysis *in vivo* of translation with nucleotide resolution using ribosome profiling. *Science* **324**, 218–223 (2009).
21. McGrew, L. L., Dworkin-Rastl, E., Dworkin, M. B. & Richter, J. D. Poly(A) elongation during *Xenopus* oocyte maturation is required for translational recruitment and is mediated by a short sequence element. *Genes Dev.* **3**, 803–815 (1989).
22. Paris, J. & Richter, J. D. Maturation-specific polyadenylation and translational control: diversity of cytoplasmic polyadenylation elements, influence of poly(A) tail size, and formation of stable polyadenylation complexes. *Mol. Cell. Biol.* **10**, 5634–5645 (1990).
23. Paris, J. & Philippe, M. Poly(A) metabolism and polysomal recruitment of maternal mRNAs during early *Xenopus* development. *Dev. Biol.* **140**, 221–224 (1990).
24. Simon, R., Tassan, J. P. & Richter, J. D. Translational control by poly(A) elongation during *Xenopus* development: differential repression and enhancement by a novel cytoplasmic polyadenylation element. *Genes Dev.* **6**, 2580–2591 (1992).
25. Vassalli, J. D. et al. Regulated polyadenylation controls mRNA translation during meiotic maturation of mouse oocytes. *Genes Dev.* **3**, 2163–2171 (1989).
26. Gebauer, F., Xu, W., Cooper, G. M. & Richter, J. D. Translational control by cytoplasmic polyadenylation of c-mos mRNA is necessary for oocyte maturation in the mouse. *EMBO J.* **13**, 5712–5720 (1994).
27. Wu, L. et al. CPEB-mediated cytoplasmic polyadenylation and the regulation of experience-dependent translation of α -CaMKII mRNA at synapses. *Neuron* **21**, 1129–1139 (1998).
28. Oh, B., Hwang, S., McLaughlin, J., Solter, D. & Knowles, B. B. Timely translation during the mouse oocyte-to-embryo transition. *Development* **127**, 3795–3803 (2000).
29. Burns, D. M. & Richter, J. D. CPEB regulation of human cellular senescence, energy metabolism, and p53 mRNA translation. *Genes Dev.* **22**, 3449–3460 (2008).
30. Novoa, I., Gallego, J., Ferreira, P. G. & Mendez, R. Mitotic cell-cycle progression is regulated by CPEB1 and CPEB4-dependent translational control. *Nature Cell Biol.* **12**, 447–456 (2010).
31. Schwanhäusser, B. et al. Global quantification of mammalian gene expression control. *Nature* **473**, 337–342 (2011).
32. Bartel, D. P. MicroRNAs: target recognition and regulatory functions. *Cell* **136**, 215–233 (2009).
33. Baek, D. et al. The impact of microRNAs on protein output. *Nature* **455**, 64–71 (2008).
34. Hendrickson, D. G. et al. Concordant regulation of translation and mRNA abundance for hundreds of targets of a human microRNA. *PLoS Biol.* **7**, e1000238 (2009).
35. Guo, H., Ingolia, N. T., Weissman, J. S. & Bartel, D. P. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* **466**, 835–840 (2010).
36. Bazzini, A. A., Lee, M. T. & Giraldez, A. J. Ribosome profiling shows that miR-430 reduces translation before causing mRNA decay in zebrafish. *Science* **336**, 233–237 (2012).
37. Braun, J. E., Huntzinger, E. & Izaurralde, E. A molecular link between miRNAs and deadenylases provides new insight into the mechanism of gene silencing by microRNAs. *Cold Spring Harb. Perspect. Biol.* **4**, a012328 (2012).
38. Audic, Y., Omilli, F. & Osborne, H. B. Postfertilization deadenylation of mRNAs in *Xenopus laevis* embryos is sufficient to cause their degradation at the blastula stage. *Mol. Cell. Biol.* **17**, 209–218 (1997).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank D. Weinberg, V. Auyeung, I. Ulitsky, C. Jan, J.-W. Nam, A. Shkumatava, S.-J. Hong, Y. Erlich and the Whitehead Genome Technology Core (V. Dhanapal, L. Francis, S. Gupta and T. Volkert) for discussions; J.-W. Nam, I. Ulitsky and D. Weinberg for assistance with transcript annotation; C. Bresilla, X. Guo, S.-J. Hong and A. Rothman for experimental assistance; and D. Weinberg for comments on the manuscript. Supported by NIH grant GM067031 (D.P.B.) and NIH Medical Scientist Training Program fellowship T32GM007753 (A.O.S.). D.P.B. is an investigator of the Howard Hughes Medical Institute.

Author Contributions A.O.S. developed PAL-seq, generated tail-length measurements, and performed associated analyses. S.W.E. performed ribosome profiling, RNA-seq and associated analyses. G.R.C. performed zebrafish injections and assisted with staging. D.P.B. supervised with help from H.S. All authors helped to design the study and write the manuscript.

Author Information Sequencing data and the processed data for each gene are available at the Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo>) under accession number GSE52809. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.P.B. (dbartel@wi.mit.edu).

A ring system detected around the Centaur (10199) Chariklo

F. Braga-Ribas¹, B. Sicardy², J. L. Ortiz³, C. Snodgrass⁴, F. Roques², R. Vieira-Martins^{1,5,6}, J. I. B. Camargo¹, M. Assafin⁵, R. Duffard³, E. Jehin⁷, J. Pollock⁸, R. Leiva⁹, M. Emilio¹⁰, D. I. Machado^{11,12}, C. Colazo^{13,14}, E. Lellouch², J. Skottfelt^{15,16}, M. Gillon⁷, N. Ligier², L. Maquet², G. Benedetti-Rossi¹, A. Ramos Gomes Jr⁵, P. Kervella², H. Monteiro¹⁷, R. Sfair¹⁸, M. El Moutamid^{2,6}, G. Tancredi^{19,20}, J. Spagnotto²¹, A. Maury²², N. Morales³, R. Gil-Hutton²³, S. Roland¹⁹, A. Ceretta^{20,24}, S.-h. Gu^{25,26}, X.-b. Wang^{25,26}, K. Harpsøe^{15,16}, M. Rabus^{9,27}, J. Manfroid⁷, C. Opitom⁷, L. Vanzi²⁸, L. Mehret¹⁰, L. Lorenzini¹¹, E. M. Schneider^{14,29,30,31}, R. Melia¹⁴, J. Lecacheux², F. Colas⁶, F. Vachier⁶, T. Widemann², L. Almenares^{19,20}, R. G. Sandness²², F. Char³², V. Perez^{19,20}, P. Lemos²⁰, N. Martinez^{19,20}, U. G. Jørgensen^{15,16}, M. Dominik³³, F. Roig¹, D. E. Reichart³⁴, A. P. LaCluyze³⁴, J. B. Haislip³⁴, K. M. Ivarsen³⁴, J. P. Moore³⁴, N. R. Frank³⁴ & D. G. Lambas^{14,30}

Hitherto, rings have been found exclusively around the four giant planets in the Solar System¹. Rings are natural laboratories in which to study dynamical processes analogous to those that take place during the formation of planetary systems and galaxies. Their presence also tells us about the origin and evolution of the body they encircle. Here we report observations of a multichord stellar occultation that revealed the presence of a ring system around (10199) Chariklo, which is a Centaur—that is, one of a class of small objects orbiting primarily between Jupiter and Neptune—with an equivalent radius of 124 ± 9 kilometres (ref. 2). There are two dense rings, with respective widths of about 7 and 3 kilometres, optical depths of 0.4 and 0.06, and orbital radii of 391 and 405 kilometres. The present orientation of the ring is consistent with an edge-on geometry in 2008, which provides a simple explanation for the dimming³ of the Chariklo system between 1997 and 2008, and for the gradual disappearance of ice and other absorption features in its spectrum over the same period^{4,5}. This implies that the rings are partly composed of water ice. They may be the remnants of a debris disk, possibly confined by embedded, kilometre-sized satellites.

Chariklo is the largest known Centaur orbiting in a region between Saturn and Uranus, and has an orbital eccentricity of 0.175 and a semi-major axis of 15.8 astronomical units (1 AU is the Earth–Sun distance). It may be a former trans-Neptunian object that has been recently (less than 10 Myr ago) scattered by gravitational perturbations from Uranus⁶. No clear detection of Chariklo's rotation has been made so far. Its surface is very dark, with a geometric albedo² of 0.035 ± 0.011 , and it is subject to long-term spectral and photometric variabilities^{3–5}, although no cometary activity has ever been reported.

An occultation of an $R = 12.4$ mag star by Chariklo was predicted⁷ to cross South America on 3 June 2013 (Extended Data Figs 1 and 2).

We obtained data from sites in Brazil, Argentina, Uruguay and Chile (Extended Data Table 1). Although the occultation by Chariklo itself was recorded at three sites in Chile, seven sites detected a total of thirteen rapid stellar flux interruptions (secondary events), two of them being resolved into two sub-events by the Danish 1.54-m telescope at the European Southern Observatory at La Silla, Chile (Fig. 1).

Displayed in the Extended Data and analysed in the Supplementary Information, all those secondary events (Extended Data Tables 2 and 3) can be readily explained by the presence of two narrow and azimuthally homogeneous rings (Fig. 2), whose widths and optical depths are given in Table 1. Even if the events were generally not resolved (Extended Data Fig. 3), their depths provide a measure of the integrated light loss of the events, which in turn depends on the local geometry of the occultation in the plane of the sky. The fact that all the events are consistent with an azimuthally homogeneous ring system makes other interpretations, such as an ensemble of cometary jets, very unlikely.

Other evidence supports our interpretation of a flat circular ring system around Chariklo. The ellipse fitted to the secondary events provides two possible ring pole positions (Table 1 and Extended Data Table 4). Our preferred solution is the one in which the rings had an opening angle of 60° in 1996–1997 and vanished from view as they were observed edge-on in 2008, owing to the orbital motion of Chariklo relative to the Earth. This provides a simple explanation for the gradual dimming of Chariklo's system, by a factor of 1.75, during that period³. Further evidence is that the 2- μm water-ice band and the spectral slope below 0.55 μm gradually disappeared^{4,5} between 1997 and 2008, implying that water ice is present in the rings. Observations made in 2013 show that the system has brightened by a factor of about 1.5 since 2008, and that the water-ice band is detectable again, supporting our interpretation (R.D. *et al.*, manuscript in preparation).

¹Observatório Nacional/MCTI, Rua General José Cristino 77, CEP 20921-400 Rio de Janeiro, RJ, Brazil. ²LESIA, Observatoire de Paris, CNRS UMR 8109, Université Pierre et Marie Curie, Université Paris-Diderot, 5 place Jules Janssen, F-92195 Meudon Cedex, France. ³Instituto de Astrofísica de Andalucía, CSIC, Apartado 3004, 18080 Granada, Spain. ⁴Max Planck Institute for Solar System Research, Justus-von-Liebig-Weg 3, 37077 Göttingen, Germany. ⁵Observatório do Valongo/UFRJ, Ladeira Pedro Antonio 43, CEP 20.080-090 Rio de Janeiro, RJ, Brazil. ⁶Observatoire de Paris, IMCCE, UPMC, CNRS, 77 Avenue Denfert-Rochereau, 75014 Paris, France. ⁷Institut d'Astrophysique de l'Université de Liège, Allée du 6 Août 17, B-4000 Liège, Belgium. ⁸Physics and Astronomy Department, Appalachian State University, Boone, North Carolina 28608, USA. ⁹Instituto de Astrofísica, Facultad de Física, Pontificia Universidad Católica de Chile, Avenida Vicuña Mackenna 4860, Santiago 7820436, Chile.

¹⁰Universidade Estadual de Ponta Grossa, O.A. - DEGE, Avenida Carlos Cavalcanti 4748, Ponta Grossa 84030-900, PR, Brazil. ¹¹Polo Astronómico Casimiro Montenegro Filho/FPTI-BR, Avenida Tancredo Neves 6731, CEP 85867-900, Foz do Iguaçu, PR, Brazil. ¹²Universidade Estadual do Oeste do Paraná (Unioeste), Avenida Tarquínio Joslin dos Santos, 1300, CEP 85870-650, Foz do Iguaçu, PR, Brazil.

¹³Ministerio de Educación de la Provincia de Córdoba, Santa Rosa 751, Córdoba 5000, Argentina. ¹⁴Observatorio Astronómico, Universidad Nacional de Córdoba, Laprida 854, Córdoba 5000, Argentina.

¹⁵Niels Bohr Institute, University of Copenhagen, Julianes Maries vej 30, 2100 Copenhagen, Denmark. ¹⁶Centre for Star and Planet Formation, Geological Museum, Øster Voldgade 5, 1350 Copenhagen, Denmark. ¹⁷Instituto de Física e Química, Avenida BPS 1303, CEP 37500-903, Itajubá, MG, Brazil. ¹⁸UNESP - Universidade Estadual Paulista, Avenida Ariberto Pereira da Cunha, 333, CEP 12516-410 Guaratinguetá, SP, Brazil. ¹⁹Observatorio Astronómico Los Molinos, DICYT, MEC, 12400 Montevideo, Uruguay. ²⁰Departamento de Astronomía, Facultad Ciencias, Universidad de la República, 11300 Montevideo, Uruguay. ²¹Observatorio El Cateajo, Mussio 255, Santa Rosa, La Pampa 6300, Argentina. ²²San Pedro de Atacama Celestial Explorations, Casilla 21, San Pedro de Atacama 1410000, Chile.

²³Complejo Astronómico El Leoncito (CASLEO) and San Juan National University, Avenida España 1512 sur, J5402DSP, San Juan, Argentina. ²⁴Observatorio del IPA, Consejo de Formación en Educación, 11800 Montevideo, Uruguay. ²⁵Yunnan Observatories, Chinese Academy of Sciences, Kunming 650011, China. ²⁶Key Laboratory for the Structure and Evolution of Celestial Objects, Chinese Academy of Sciences, Kunming 650011, China. ²⁷Max Planck Institute for Astronomy, Königstuhl 17, 69117 Heidelberg, Germany. ²⁸Department of Electrical Engineering and Center of Astro-Engineering, Pontificia Universidad Católica de Chile, Avenida Vicuña Mackenna 4860, Santiago 7820436, Chile. ²⁹Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Córdoba 5000, Argentina. ³⁰Instituto de Astronomía Teórica y Experimental IATE-CONICET, Córdoba 5000, Argentina. ³¹Facultad de Ciencias Exactas, Físicas y Naturales, Universidad Nacional de Córdoba (UNC), Córdoba 5000, Argentina.

³²Unidad de Astronomía, Facultad de Ciencias Básicas, Universidad de Antofagasta, Avenida Angamos 601, Antofagasta, Region II, Chile. ³³Scottish Universities Physics Alliance, University of St Andrews, School of Physics and Astronomy, North Haugh, St Andrews KY16 9SS, UK. ³⁴Department of Physics and Astronomy, University of North Carolina - Chapel Hill, North Carolina 27599, USA.

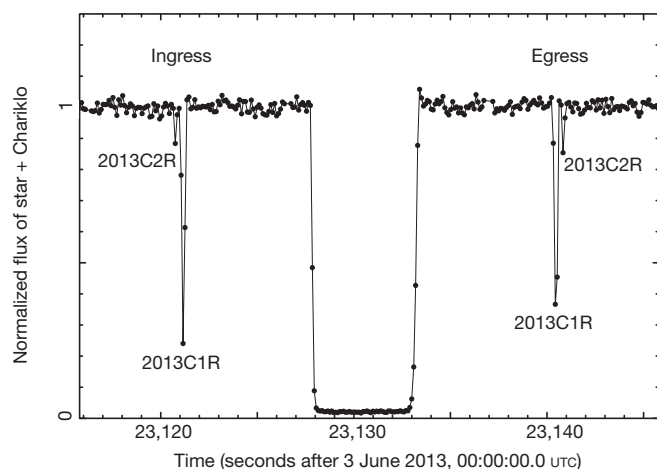


Figure 1 | Light curve of the occultation by the Chariklo system. The data were taken with the Danish 1.54-m telescope (La Silla) on 3 June 2013, at a rate of almost 10 Hz and with a long-pass filter and a cut-off below 650 nm, limited at long wavelengths by the sensitivity of the charge-coupled-device chip (Supplementary Information). Aperture photometry provided the flux from the target star and a fainter nearby reference star. Low-frequency sky transparency variations were removed by dividing the target flux by an optimal running average of 87 data points (8.7 s) of the reference star, resulting in a final signal-to-noise ratio of 64 per data point. The sum of the stellar and Chariklo fluxes has been normalized to unity outside the occultation. The central drop is caused by Chariklo, and two secondary events, 2013C1R and 2013C2R, are observed, one at ingress (before the main Chariklo occultation) and then at egress (after the main occultation). A more detailed view of these ring events is shown in Fig. 3.

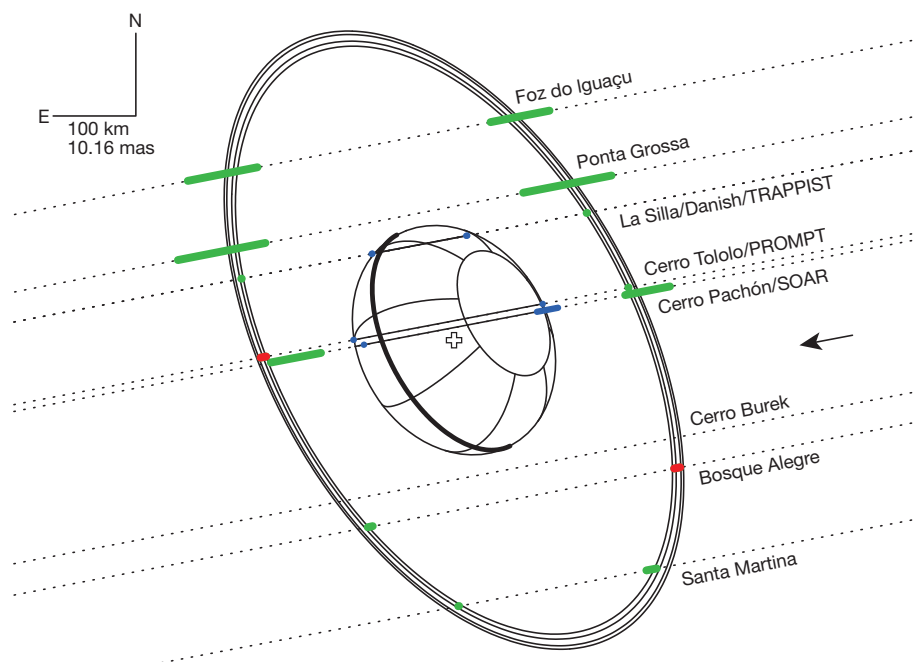


Figure 2 | Chariklo ring system. The dotted lines are the trajectories of the star relative to Chariklo in the plane of the sky, as observed from eight sites (Supplementary Information), the arrow indicating the direction of motion. The green segments represent the locations of ring C1R observed at each station (1σ uncertainty). For clarity, we have not plotted the detections made at the TRAPPIST and 0.275-m telescopes (at La Silla and Bosque Alegre, respectively) because they have larger error bars than their local counterparts, and would supersede the corresponding green segments. Two ring events occurred during camera readout times (red segments) at Bosque Alegre and Cerro Tololo, and also provide constraints on the ring orbit. The ring events are only

Owing to its higher acquisition rate (10 Hz), the Lucky Imager camera^{8,9} of the Danish 1.54-m telescope actually resolved the secondary events into two rings, denoted 2013C1R and 2013C2R (C1R and C2R for short) in Figs 1 and 3. We use the terms ingress and egress refer to the first and, respectively, second of a pair of ring events at a given site. All the Danish events are satisfactorily fitted by sharp-edged ring models whose radial widths (W) in the ring plane and normal optical depths (τ_N) are listed in Table 1. We also provide the equivalent depths ($E_e = W\tau_N$), which can be related to the amount of material contained in the ring¹⁰. C2R is about 40% narrower than C1R, and contains about 12 times less material. We note that no material is detected in the gap between C1R and C2R, up to a limit of 0.004 in normal optical depth and 0.05 km in equivalent depth (Table 1 and Extended Data Table 4).

By analogy with Saturn's A ring¹¹ or the dense rings of Uranus¹², we estimate that the surface density of C1R lies in the range 30–100 g cm⁻² (Supplementary Information). Then, the mass of C1R is equivalent to that of an icy body with a radius of roughly 1 km, whereas C2R corresponds to a body of half that size. If the photometric variability of Chariklo's system between 1997 and 2008 is entirely due to the ring changing geometry³, we estimate the ring reflectivity to be $I/F \approx 0.09 \pm 0.04$ (I is the intensity of light reflected by the surface and πF is the incident solar flux density). Thus, Chariklo's ring particles would be significantly brighter than those of Uranus's rings¹³ ($I/F \approx 0.05$), but would be significantly darker than those of Saturn's A ring¹⁴ ($I/F \approx 0.3$). We note that, if part of the photometric variability is caused by Chariklo itself, then the ring material would be darker than estimated above (Supplementary Information).

Constraints on Chariklo's limb shape are based on only two occultation chords (Supplementary Information and Extended Data Table 5). Our simplest model describes an oblate Chariklo surrounded by a circular equatorial ring system (Extended Data Table 6). The fitted limb (Fig. 2) has an equivalent radius of 127 km (the radius of an equivalent

marginally detected at Cerro Burek, but the signal-to-noise ratio is not sufficient to put further constraints on the ring orbit and equivalent width. An elliptical fit to the green and red segments (excluding, because of timing problems (Supplementary Information), the SOAR events at Cerro Pachón) provides the centre of the rings (open cross), as well as their sizes, opening angle and orientation (Table 1). Chariklo's limb has been fitted to the two chords' extremities (blue segments) obtained at La Silla and Cerro Tololo, assuming that the centres of Chariklo and the rings, as well as their position angles, coincide. This is expected if Chariklo is a spheroid, with a circular ring orbiting in the equatorial plane (see text and Supplementary Information).

Table 1 | Ring physical parameters

	Radial width, W (km)	Normal optical depth, τ_N	Equivalent depth, $E_r = W\tau_N$ (km)
Ring C1R, Danish ingress	6.16 ± 0.11	0.449 ± 0.009	2.77 ± 0.04
Ring C1R, Danish egress	7.17 ± 0.14	0.317 ± 0.008	2.28 ± 0.03
Ring C2R, Danish ingress	$3.6^{+1.3}_{-2.0}$	$0.05^{+0.06}_{-0.01}$	0.18 ± 0.03
Ring C2R, Danish egress	$3.4^{+1.1}_{-1.4}$	$0.07^{+0.05}_{-0.03}$	0.24 ± 0.02
Ring C1R radius (km)		390.6 ± 3.3	
Ring C2R radius (km)		404.8 ± 3.3	
Radial separation of rings C1R and C2R (km)		14.2 ± 0.2	
Gap between rings C1R and C2R (km)		8.7 ± 0.4 ($E_r < 0.05$ km, $\tau_N < 0.004$)	
Opening angle, B (°)		$+33.77 \pm 0.41$	
Position angle, P (°)		-61.54 ± 0.14 (preferred) or 118.46 ± 0.14	
Pole position (equatorial J2000)			
	Solution 1 (preferred)	Solution 2	
Right ascension, α_p	10 h 05 min \pm 02 min	01 h 48 min \pm 01 min	
Declination, δ_p	$41^\circ 29' \pm 13'$	$03^\circ 26' \pm 19'$	

The geocentric distance of Chariklo's system at the moment of the occultation, $D = 2.031 \times 10^9$ km, provides a scale of $9,846$ km arcsec $^{-1}$ in the plane of the sky. All the secondary events apart from those of the Danish 1.54-m telescope (Fig. 3) are satisfactorily fitted by a model where the rings C1R and C2R have widths and optical depths that are averages of the Danish ingress and egress values given above, that is, $W_{C1R} = 6.6$ km, $\tau_{N,C1R} = 0.38$, $W_{C2R} = 3.4$ km and $\tau_{N,C2R} = 0.06$ (Supplementary Information). The ring opening angle, B , is the absolute value of the elevation of the observer above the ring plane. The position angle, P , is the angle between celestial north and the semi-minor axis of the ring projected in the plane of the sky, counted positively from celestial north to celestial east. By convention, it refers to the projected semi-major axis that corresponds to superior conjunction. The solution that best explains the photometric and spectral variations of Chariklo's system^{3–5} is chosen as preferred (see text).

spherical body; see Supplementary Information), consistent with the value derived from thermal data², 124 ± 9 km, thus supporting our model.

From the Roche critical density limit¹⁵, we estimate that typical densities for Chariklo and ring particles are consistent with the present ring locations (Supplementary Information). Moreover, an unperturbed ring several kilometres in width and of thickness h should spread, owing to interparticle collisions¹⁶, in $10^4/h^2$ years (Supplementary Information), or a few thousand years, assuming h of a few metres (by analogy with Saturn's rings). Furthermore, Poynting–Robertson drag¹⁶ should spread sub-centimetre particles in a few million years at most (Supplementary Information). Thus, the rings are either very young or actively confined. A confinement mechanism may be provided by kilometre-sized 'shepherd

satellites' that would have a mass comparable to that of the rings (Supplementary Information).

We do not know if rings around minor bodies stem from a generic, yet unknown, process, or are exceptional features. We note that many stellar occultations by main-belt asteroids and more than ten trans-Neptunian events^{17–20} have not revealed rings so far (nor have direct images). Stellar occultations and appulses involving (2060) Chiron (a Centaur similar in size to Chariklo) in 1993 and 1994 revealed a narrow, jet-like feature and diffuse material around that object^{21,22}. This was interpreted as material ejected from the surface, partly on the basis that Chiron is known to be an active, comet-like object. It is unclear whether the detection of material around both objects is a mere coincidence, or whether they share a common physical process (noting that no cometary activity has been detected around Chariklo).

About 5% of the Centaur and trans-Neptunian population²³ are known to have satellites. Although the large satellites are thought to result from three-body captures, their small counterparts are more likely to form from impacts²⁴, or rotational disruptions²⁵, and possibly re-accretion from a disk remaining after that event. So far, no observations have shown satellites around Chariklo (the rings span at most $0.04''$ around the primary object, making direct detections of associated small satellites a challenge). Several origins for Chariklo's rings can be proposed, all relying on a debris disk in which the largest fragments acted as shepherds for the smaller material. The first possibility is that an impactor excavated icy material from Chariklo's outer layers, destroyed a pre-existing satellite or was itself disrupted during the impact. The second is that a debris disk formed from a rotational disruption of the main body or was fed by cometary-like activity. Third, two pre-existing satellites

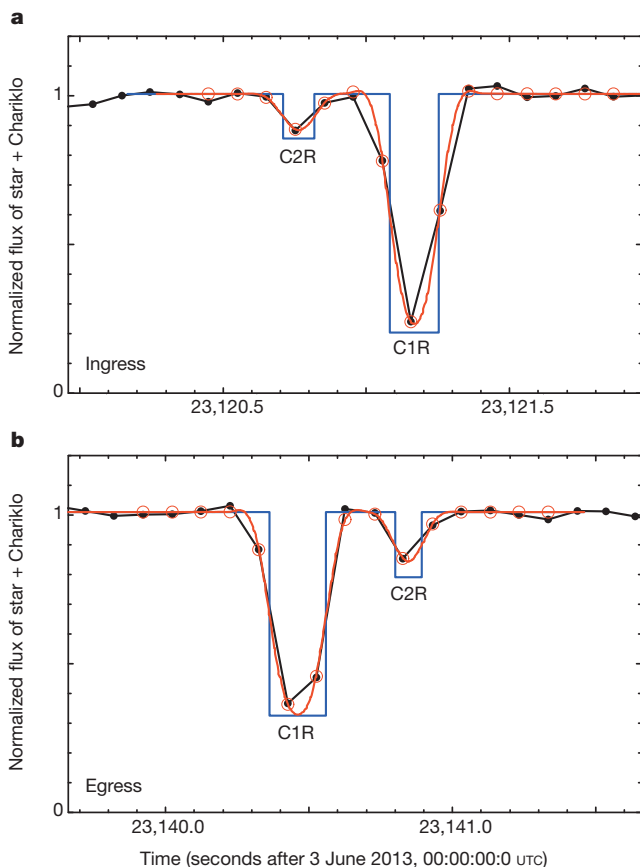


Figure 3 | Fits to the Danish ring events. **a, b,** The red curves are synthetic occultation profiles produced by semi-transparent bands with square-well profiles (the blue lines), after convolution by Fresnel diffraction, observed bandwidth, the stellar radius projected at Chariklo, and the finite integration time. The open red circles are the values of the model for the times corresponding to the observed data points (black points) at the ingress (**a**) and egress (**b**). The χ^2 values per degree of freedom of the fits to the four ring events vary from 0.4 to 1.2 (Extended Data Table 2). This indicates satisfactory fits, and shows that the events are compatible with sharp-edged rings. The resulting widths and optical depths of rings C1R and C2R are listed in Table 1, after the appropriate projections into the plane of the rings have been performed. Extended Data Table 3 shows that the widths and optical depths of C1R at the Danish 1.54-m telescope differ moderately but significantly between ingress and egress. The equivalent depth of C1R changes by 21% between ingress and egress. Similar variations are observed in Uranus's narrow rings, and might be associated with normal mode oscillations that azimuthally modulate the width and optical depth of the rings¹⁰. Differences between C2R ingress and egress are marginally significant.

might have collided through a mechanism yet to be explained. Finally, a retrograde satellite might have migrated inwards and eventually been disrupted by tidal forces.

We note that the mass and angular momentum of the rings and their hypothetical shepherds are very small (by a factor of less than 10^{-5}) compared with that of Chariklo. The typical escape velocity at the surface of Chariklo is $\sim 0.1 \text{ km s}^{-1}$. Thus, if an impact from an outsider generated the rings, it must have struck at low velocity. Whereas the impact velocities in the main belt of asteroids are of the order of 5 km s^{-1} , they are $\sim 1 \text{ km s}^{-1}$ in the outer Solar System, and were even lower before the Kuiper Belt was dynamically excited²⁶, which may explain why no rings have yet been found around main-belt asteroids. Finally, Chariklo's orbit is perturbed by Uranus, which transferred the Centaur from the trans-Neptunian region less than $\sim 10 \text{ Myr}$ ago⁶. As estimated in Supplementary Information, a very close encounter at about five Uranus radii is actually necessary to disrupt the ring system. Such an event has a small probability of occurrence²⁷, which supports the possibility that the rings formed in the trans-Neptunian region and survived the transfer episode.

Received 23 December 2013; accepted 11 February 2014.

Published online 26 March 2014.

1. Tiscareno, M. S. in *Planets, Stars and Stellar Systems* Vol. 3 *Solar and Stellar Planetary Systems* (eds Oswalt, T. D., French, L. M. & Kalas, P.) 309–375 (Springer, 2013).
2. Fornasier, S. *et al.* TNOs are cool: A survey of the trans-Neptunian region VIII. Combined Herschel PACS and SPIRE observations of nine bright targets at 70–500 μm . *Astron. Astrophys.* **555**, A15 (2013).
3. Belskaya, I. N. *et al.* Polarimetry of Centaurs (2060) Chiron, (5145) Pholus and (10199) Chariklo. *Icarus* **210**, 472–479 (2010).
4. Guilbert, A. *et al.* A portrait of Centaur 10199 Chariklo. *Astron. Astrophys.* **501**, 777–784 (2009).
5. Guilbert-Lepoutre, A. A thermal evolution model of Centaur 10199 Chariklo. *Astron. J.* **141**, 103 (2011).
6. Horner, J., Evans, N. W. & Bailey, M. E. Simulations of the population of Centaurs - I. The bulk statistics. *Mon. Not. R. Astron. Soc.* **354**, 798–810 (2004).
7. Camargo, J. I. B. *et al.* Candidate stellar occultations by Centaurs and trans-Neptunian objects up to 2014. *Astron. Astrophys.* **561**, A37 (2014).
8. Harpsøe, K. B. W., Jørgensen, U. G., Andersen, M. I. & Grundahl, F. High frame rate imaging based photometry. Photometric reduction of data from electron-multiplying charge coupled devices (EMCCDs). *Astron. Astrophys.* **542**, A23 (2012).
9. Skottfelt, J. *et al.* EMCCD photometry reveals two new variable stars in the crowded central region of the globular cluster NGC 6981. *Astron. Astrophys.* **553**, A111 (2013).
10. French, R. G., Nicholson, P. D., Porco, C. C. & Marouf, E. A. in *Uranus* (eds Bergstrahl, J. T., Miner, E. D. & Matthews, M. S.) 327–409 (Univ. Arizona Press, 1991).
11. Colwell, J. E. *et al.* in *Saturn from Cassini-Huygens* (eds Dougherty, M. K., Esposito, L. W. & Krimigis, S. M.) 375–412 (Springer, 2009).
12. Esposito, L. W., Brahic, A., Burns, J. A. & Marouf, E. A. in *Uranus* (eds Bergstrahl, J. T., Miner, E. D. & Matthews, M. S.) 410–465 (Univ. Arizona Press, 1991).
13. Karkoschka, E. Comprehensive photometry of the rings and 16 satellites of Uranus with the Hubble Space Telescope. *Icarus* **151**, 51–68 (2001).
14. Hedman, M. M. *et al.* Connections between spectra and structure in Saturn's main rings based on Cassini VIMS data. *Icarus* **223**, 105–130 (2013).
15. Tiscareno, M. S., Hedman, M. M., Burns, J. A. & Castillo-Rogez, J. Compositions and origins of outer planet systems: insights from the Roche critical density. *Astrophys. J.* **765**, L28 (2013).
16. Goldreich, P. & Tremaine, S. Towards a theory for the Uranian rings. *Nature* **277**, 97–99 (1979).
17. Elliot, J. L. *et al.* Size and albedo of Kuiper belt object 55636 from a stellar occultation. *Nature* **465**, 897–900 (2010).
18. Sicardy, B. *et al.* A Pluto-like radius and a high albedo for the dwarf planet Eris from an occultation. *Nature* **478**, 493–496 (2011).
19. Ortiz, J. L. *et al.* Albedo and atmospheric constraints of dwarf planet Makemake from a stellar occultation. *Nature* **491**, 566–569 (2012).
20. Braga-Ribas, F. *et al.* The size, shape, albedo, density, and atmospheric limit of Transneptunian object (50000) Quaoar from multi-chord stellar occultations. *Astrophys. J.* **773**, 26 (2013).
21. Bus, S. J. *et al.* Stellar occultation by 2060 Chiron. *Icarus* **123**, 478–490 (1996).
22. Elliot, J. L. *et al.* Jet-like features near the nucleus of Chiron. *Nature* **373**, 46–49 (1995).
23. Brunini, A. On the dynamical evolution and end states of binary Centaurs. *Mon. Not. R. Astron. Soc.* **437**, 2297–2302 (2014).
24. Noll, K. S., Grundy, W. M., Chiang, E. I., Margot, J.-L. & Kern, S. D. in *The Solar System Beyond Neptune* (eds Barucci, M. A., Boehnhardt, H., Cruikshank, D. P. & Morbidelli, A.) 345–363 (Univ. Arizona Press, 2008).
25. Ortiz, J. L. *et al.* Rotational fission of trans-Neptunian objects: the case of Haumea. *Mon. Not. R. Astron. Soc.* **419**, 2315–2324 (2012).
26. Ćuk, M., Ragozzine, D. & Nesvorný, D. On the dynamics and origin of Haumea's moons. *Astron. J.* **146**, 89 (2013).
27. Nogueira, E., Brasser, R. & Gomes, R. Reassessing the origin of Triton. *Icarus* **214**, 113–130 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank S. Fornasier, I. Belskaya, B. Carry, E. Nogueira, P. Michel and A. Morbidelli for the discussions that helped to interpret our results. We also thank J. A. Burns for comments that helped to improve the paper. F.B.-R. acknowledges LineA (Laboratório Interinstitucional de e-Astronomia) for hosting the campaign webpage and the support (grant 150541/2013-9) of CNPq, Brazil. Operation of the Danish 1.54-m telescope is financed by a grant to U.G.J. by the Danish Natural Science Research Council (FNU) and by the Centre for Star and Planet Formation (StarPlan). F. Colas, J.L., L. Maquet, F. Roques, B.S., F.V. and T.W. acknowledge support from the French grant 'Beyond Neptune II'. J.L.O., R.D. and N. Morales acknowledge funding from Spanish AYA grants and FEDER funds. S.-h.G. and X.-b.W. are grateful for financial support from the National Natural Science Foundation of China through grants 10873031 and 11073051. M.D. is a Royal Society University Research Fellow at St Andrews University. TRAPPIST is a project funded by the Belgian Fund for Scientific Research (FRS-FNRS) with the participation of the Swiss National Science Foundation (SNF). E.J. and M.G. are FNRS Research Associates, J.M. is FNRS Research Director. C.O. thanks the Belgian FNRS for funding her PhD thesis. J.I.B.C., M.A. and R.V.-M. acknowledge CNPq grants 302657/2010-0, 482080/2009-4 478318/2007-3 and 304124/2007-9, respectively. The Universidad Católica Observatory (UCO) Santa Martina is operated by the Pontificia Universidad Católica de Chile (PUC). C.S. received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 268421. M.R. acknowledges support from FONDECYT postdoctoral fellowship 3120097. This work is partly based on observations made at the Pico dos Dias Observatory from the National Laboratory of Astrophysics (OPD/LNA) – Brazil; and partly based on observations obtained at the Southern Astrophysical Research (SOAR) telescope, which is a joint project of the Ministério da Ciência, Tecnologia, e Inovação (MCT) da República Federativa do Brasil, the US National Optical Astronomy Observatory (NOAO), the University of North Carolina at Chapel Hill (UNC-CH) and Michigan State University (MSU). This publication makes use of data products from the Two Micron All Sky Survey, a joint project of the University of Massachusetts and the Infrared Processing and Analysis Center/California Institute of Technology, funded by the National Aeronautics and Space Administration (NASA) and the National Science Foundation (NSF). It also makes use of data products from the Wide-field Infrared Survey Explorer, which is a joint project of the University of California, Los Angeles, and the Jet Propulsion Laboratory/California Institute of Technology, funded by NASA. UNC-CH gratefully acknowledges NSF awards 0959447, 1009052 and 1211782 for support of Skynet/PROMPT. L.V. and R.L. acknowledge support by CONICYT through the project Anillo ACT-86. A.M. acknowledges the use of C. Harlinton's 20-inch Planewave telescope, which is part of the Searchlight Observatory Network.

Author Contributions F.B.-R. planned the observation campaign, centralized the occultation predictions, participated in the observations, analysed data and results, ran the diffraction and limb-fitting codes, and wrote the paper. B.S. helped to plan the campaign, analysed data and results, wrote and ran the diffraction and limb-fitting codes, and wrote the paper. J.L.O. helped to plan the campaign, analysed data for the occultation prediction, and obtained and analysed data and results. C.S. coordinated and analysed the data from Danish 1.54-m telescope. F. Roques analysed data, and wrote and ran the diffraction and limb-fitting codes. R.V.-M. coordinated the predictions and analysed data. J.I.B.C., F.B.-R., R.V.-M. and M.A. discovered the star candidate and analysed data for the occultation prediction. R.D. coordinated the SOAR and Bosque Alegre observations, and helped to analyse the results. E.J., J.P., R.L., M.E., D.I.M., C.C. and J. Spagnotto obtained and analysed the positive occultation detections from the TRAPPIST, PROMPT, UCO Santa Martina, UEPG, Foz do Iguaçu, Bosque Alegre and Danish telescopes, respectively. E.L. helped to analyse the results. J. Skottfelt, M.G., L. Maquet, N.L., G.B.-R. and A.R.G. helped to obtain the occultation light curves or analysed the data, or both. P.K. ran stellar models to determine the apparent star diameter. H.M. obtained the stellar spectrum. R.S. calculated the short-term ring stability. M.E.M. worked on the confinement mechanisms of the rings. G.T., J.P., A.M., N. Morales, R.G.-H., S.R. and A.C. obtained the negative detections. S.-h.G., X.-b.W., K.H., M.R., J.M., C.O., L.V., L. Mehret, L.L., E.M.S. and R.M. helped to obtain and analyse the positive occultation detections. J.L., F. Colas, F.V., T.W. and K.H. helped to make the predictions and analyse the data. L.A., R.G.S., F. Char, V.P., P.L., N. Martinez, U.G.J., M.D., F. Roig, D.E.R., A.P.L., J.B.H., K.M.I., J.P.M., N.R.F. and D.G.L. were involved in event observations. All authors were given the opportunity to review the results and comment on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to F.B.-R. (ribas@on.br).

Efficient rotational cooling of Coulomb-crystallized molecular ions by a helium buffer gas

A. K. Hansen¹, O. O. Versolato², L. Klosowski³, S. B. Kristensen¹, A. Gingell¹, M. Schwarz², A. Windberger², J. Ullrich^{2,4}, J. R. Crespo López-Urrutia² & M. Drewsen¹

The preparation of cold molecules is of great importance in many contexts, such as fundamental physics investigations^{1,2}, high-resolution spectroscopy of complex molecules^{3–5}, cold chemistry^{6,7} and astrochemistry⁸. One versatile and widely applied method to cool molecules is helium buffer-gas cooling in either a supersonic beam expansion^{9,10} or a cryogenic trap environment^{11,12}. Another more recent method applicable to trapped molecular ions relies on sympathetic translational cooling, through collisional interactions with co-trapped, laser-cooled atomic ions, into spatially ordered structures called Coulomb crystals, combined with laser-controlled internal-state preparation^{6,7,13–23}. Here we present experimental results on helium buffer-gas cooling of the rotational degrees of freedom of MgH^+ molecular ions, which have been trapped and sympathetically cooled¹³ in a cryogenic linear radio-frequency quadrupole trap. With helium collision rates of only about ten per second—that is, four to five orders of magnitude lower than in typical buffer-gas cooling settings—we have cooled a single molecular ion to a rotational temperature of $7.5^{+0.9}_{-0.7}$ kelvin, the lowest such temperature so far measured. In addition, by varying the shape of, or the number of atomic and molecular ions in, larger Coulomb crystals, or both, we have tuned the effective rotational temperature from about 7 kelvin to about 60 kelvin by changing the translational micromotion energy of the ions²⁴. The extremely low helium collision rate may allow for sympathetic sideband cooling of single molecular ions, and eventually make quantum-logic spectroscopy²⁵ of buffer-gas-cooled molecular ions feasible. Furthermore, application of the present cooling scheme to complex molecular ions should enable single- or few-state manipulations of individual molecules of biological interest^{4,5}.

Cooling of molecules through collisional interactions with a cryogenically cooled He buffer gas has in the past decades been applied to produce both translationally and internally cold neutral^{7,12} as well as ionic^{3,11,26} molecular species. For trapped ions, such experiments have been conducted in various types of radio-frequency traps, with the lowest temperatures, of ~ 10 – 20 K, reached in multipole traps^{11,26}. In such experiments, translational, vibrational and rotational degrees of freedom have to be cooled simultaneously, with the result that He densities of the order of 10^{14} – 10^{15} cm^{-3} are required to achieve efficient cooling.

In recent years, researchers have investigated alternative strategies to produce cold molecular ions. These have been based on combining translational sympathetic cooling through interactions with directly laser-cooled atomic ions^{6,7,14–23} with laser-based methods for internal-state preparation^{18–20}. In these scenarios, the translational-cooling scheme is indeed very versatile and can fairly easily bring the translational temperature into the millikelvin¹³ and even the microkelvin range, although internal-state preparation has to be tailored to the specific molecule^{18–20}. The ideal state preparation scheme would, however, be one that provides extremely low translational temperatures and cold internal-state distributions simultaneously, and, in addition, allows for coherent manipulation of the molecules.

In the present Letter, we present results of He buffer-gas cooling in a novel setting where MgH^+ ions are trapped in a cryogenically cooled, linear, radio-frequency quadrupole trap and translationally cooled through Coulomb interaction with simultaneously trapped, laser-cooled atomic Mg^+ ions¹⁸. Under these conditions, the interaction with the He buffer gas is needed only to cool the rovibrational degrees of freedom, and, consequently, He densities of only $\sim 10^{10}$ cm^{-3} , that is, four to five orders of magnitude lower than in typical buffer-gas cooling settings, are sufficient for efficient cooling. The internal-state cooling rate has simply to exceed any heating rates present in the trapping environment. Because the vibrational degree of freedom of the MgH^+ ions is already frozen out at room temperature ($>99.9\%$ probability of being in the vibrational ground state), at the cryogenic temperatures considered here this degree of freedom can be entirely disregarded. In contrast to previous experiments on He buffer-gas cooling, we directly measure the full rotational-state distribution of the cold molecules. For a single MgH^+ ion, we obtain a rotational-state distribution corresponding to a temperature of $7.5^{+0.9}_{-0.7}$ K, the lowest yet reported for a buffer-gas-cooled molecular ion. Furthermore, by exploiting the fact that ions positioned away from the radio-frequency field-free axis of the linear trap exhibit a driven motion at the trapping radio frequency, called ‘micromotion’²⁴, we are able to tune the effective collisional temperature between the MgH^+ ions and He atoms from ~ 7 K to ~ 60 K, just by changing the shape, size or ion composition of the Coulomb crystal. Importantly, we prove that the rotational temperature is in equilibrium with the effective collisional temperature. These results have immediate ramifications for studies of state-dependent processes involving smaller molecular ions of astrophysical interest. Furthermore, they should make it possible to perform single-ion experiments with larger molecular ions of biological relevance under significantly improved control.

The apparatus used in the experiments is described in detail elsewhere^{23,27}, and the essential parameters for calculating the effect of micromotion on the collisional dynamics are presented in Supplementary Information. A typical experimental sequence starts with isotope-selective loading of $^{24}\text{Mg}^+$ ions into the trap through resonance-enhanced photoionization of atoms from an effusive beam crossing the centre of the trap²³. By exciting the $3s\ ^2S_{1/2} \rightarrow 3p\ ^2P_{3/2}$ transition of these ions using light of wavelength 280 nm, they are laser-cooled into a Coulomb crystal. Next, H_2 gas is leaked into the inner trap enclosure (kept at ~ 5 K) through a thin stainless steel tube in thermal contact with a surrounding thermal shielding box at ~ 40 K, and MgH^+ ions are produced in exothermic reactions with Mg^+ ions in the $3p\ ^2P_{3/2}$ state¹³. When the desired fraction of atomic ions has reacted, a valve controlling the H_2 gas inlet is closed and He gas is leaked in through the same tube. Within a fraction of a second (see below), the rotational-state distribution of the MgH^+ ions in the vibrational ground state equilibrates under the specified collisional conditions. Finally, through the application of resonance-enhanced multiphoton dissociation (REMPD) with rotational-state resolution, we determine the population in the individual rotational states (this technique is described in detail in ref. 18).

¹The Danish National Research Foundation Center for Quantum Optics – QUANTOP, Department of Physics and Astronomy, Aarhus University, DK-8000 Aarhus C, Denmark. ²Max-Planck-Institut für Kernphysik, Saupfercheckweg 1, D-69117 Heidelberg, Germany. ³Institute of Physics, Faculty of Physics, Astronomy and Informatics, Nicolaus Copernicus University, Grudziadzka 5, 87-100 Toruń, Poland. ⁴Physikalisch-Technische Bundesanstalt, Bundesallee 100, D-38116 Braunschweig, Germany.

In Fig. 1, projection images of a series of Coulomb crystals, in which MgH^+ ions experience various buffer-gas collision conditions due to the spatially dependent, radio-frequency-field-induced micromotion, are presented (upper panels) together with the respective measured rotational-state distributions (middle panels). It is clear in all five cases that the rotational distributions closely resemble thermal ones at different temperatures. Because the full three-dimensional velocity distribution of the MgH^+ ions due to micromotion is well known for any given crystal (Supplementary Information), knowledge of the velocity distribution of the He buffer gas in the trapping region will completely characterize the collision conditions. Fortunately, because a single MgH^+ ion sympathetically cooled by a single atomic ion as in Fig. 1a does not exhibit appreciable micromotion (the micromotion energy has been measured to be well below an equivalent temperature of 1 K), only the buffer-gas velocity distribution contributes to the collisional dynamics here. Consequently, if we assume thermal equilibrium between the translational and rotational degrees of freedom in the collision, we can infer the He gas temperature from the measured rotational-state distribution. Using this distribution, we arrive at a He gas temperature of $8.7^{+1.0}_{-0.7}$ K (Supplementary Information). The corresponding single-ion collisional-speed distribution in the centre-of-mass (COM) frame is presented in the lower panel of Fig. 1a. Using a He temperature of 8.7 K for each of the other four cases in Fig. 1, we present the calculated speed distributions in the COM frame as solid curves in the lower panels together with thermal distributions with the same average kinetic energy (dashed lines). Although the micromotion velocity distribution is not thermal, the speed distribution is still very close to being thermal. This is because micromotion is induced only along a single spatial direction at any given point in space, whereas the He velocity distribution can be assumed isotropic.

In addition to the full rotational distributions presented in Fig. 1, we have measured the rotational ground-state population for a larger set

of Coulomb crystals. In Fig. 2a, this population is plotted as a function of the average micromotion energy, E_{avg} , and is seen to be a monotonically decreasing function of E_{avg} . Larger error bars for the black points result from an uncertainty in the dissociation efficiency when not measuring the whole rotational distribution. In the inset of Fig. 2a, we plot the dissociation efficiency for the measurements with full determination of the rotational distribution, together with a conservative estimate of its range for the remaining cases at higher values of E_{avg} . Assuming rotational ground-state populations corresponding to those of thermal distributions, we show the inferred rotational temperature versus E_{avg} in Fig. 2b. Moreover, in the inset in Fig. 2b we plot the calculated COM collisional-speed distribution for the experiment with the highest value of E_{avg} . This distribution is quite distorted relative to a thermal one with the same average energy. In Fig. 3, we finally plot the rotational collisional temperature presented in Fig. 2 as a function of the average collisional temperature, T_{coll} , defined as

$$T_{\text{coll}} = \frac{2}{3k_B} E_{\text{coll}}$$

with

$$E_{\text{coll}} = \frac{75}{58} k_B T_{\text{He}} + \frac{4}{29} E_{\text{avg}}$$

where k_B is Boltzmann's constant and T_{He} is the He temperature. Details about the latter formula can be found in Supplementary Information. Within the uncertainties, we find excellent agreement between the measured rotational and the calculated collisional temperatures, for all data points. The only a-priori unknown parameter for the collisional dynamics, T_{He} , we fixed from the MgH^+ single-ion experiment to be 8.7 K. Importantly, independent control of the shape, size and composition of the ion crystal makes us able to exploit micromotion to tune the rotational

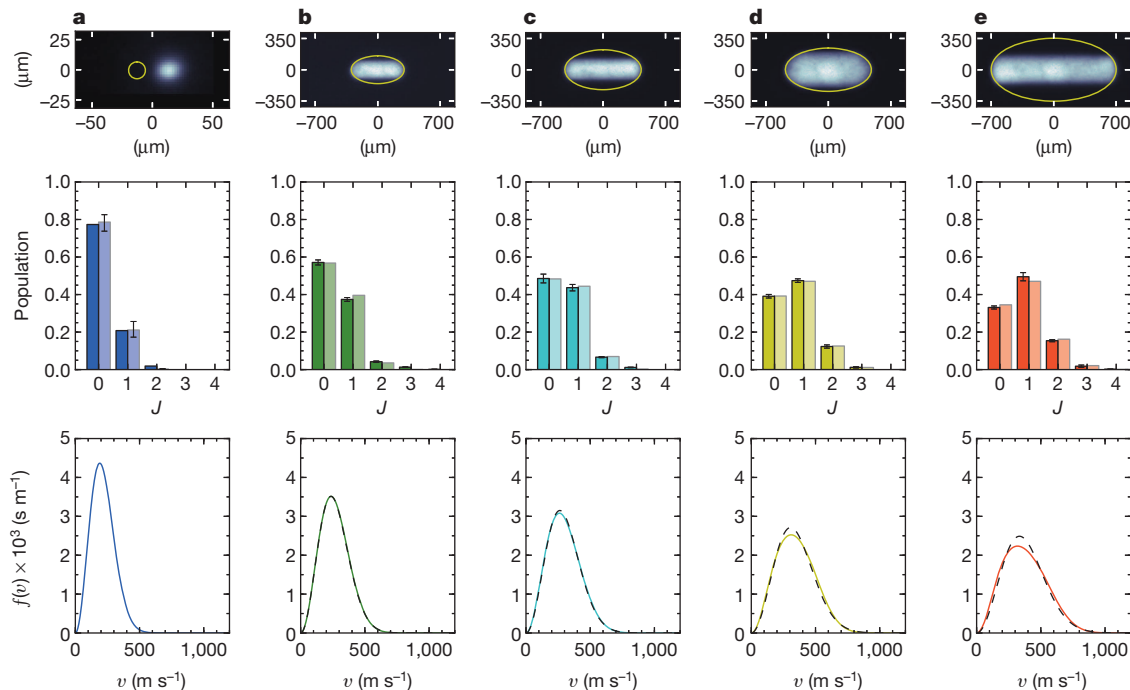


Figure 1 | Spatial, rotational and collisional speed distributions. Upper panels: projection images of a series of Coulomb crystals with a varying numbers of Mg^+ and MgH^+ ions: $(N_{\text{Mg}}, N_{\text{MgH}}) = (1, 1)$ (a), (950, 600) (b), (1,750, 2,750) (c), (4,900, 1,250) (d), (5,500, 13,200) (e). In b–e the yellow ellipse indicates the outer boundary of the Coulomb crystals, and in a the circle indicates the position of the single MgH^+ ion. The crystals are cylindrically symmetric around the horizontal axis. Only the fluorescing Mg^+ ions are visible (white regions). Centre panels: measured rotational-state distributions including 1-s.e.m. error bars (dark-coloured bars), as well as best fits to thermal distributions (light-coloured bars) for the crystals b–e. J , rotational quantum

number. For the single MgH^+ ion (a), the heights of the dark-coloured bars are obtained from the number of REMPD-induced dissociation events and the number of applied REMPD pulses. The total number of dissociation attempts was 665. The corresponding best fit of the rotational temperature is obtained from a maximum-likelihood estimate (light-coloured bars). The error bars indicate the 1-s.d. uncertainty of this estimate. The estimated rotational temperatures are 7.6 K (a), 12 K (b), 15 K (c), 20 K (d) and 23 K (e). Lower panels: calculated collisional-speed distributions in the COM reference frame of the MgH^+ –He system, for a He gas at $T_{\text{He}} = 8.7$ K (solid lines), together with thermal distributions with the same kinetic energies (dashed lines).

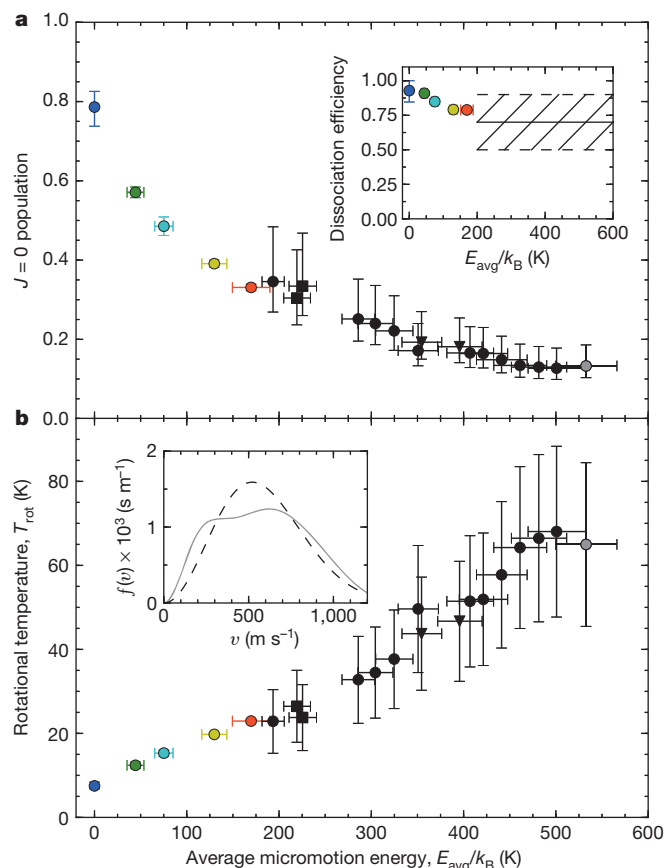


Figure 2 | Rotational ground-state populations and temperatures.

a, Measured rotational ground-state population versus average micromotion energy (see text and Supplementary Information). Error bars, 1 s.d. Inset: measured dissociation efficiencies for the experiments presented in Fig. 1 and estimations of those for the experiments where only dissociation out of the $J=0$ state was carried out. The dashed lines indicate conservative boundaries of the latter. **b**, Rotational temperatures corresponding to the measured $J=0$ populations. The coloured points correspond to the fitted distributions in Fig. 1. Error bars, 1 s.d. Inset: calculated collisional-speed distribution for the rightmost, grey-coloured, data point. For this extreme micromotion energy, a strong deviation from a thermal distribution is noticeable. The different black symbols correspond to different applied radio-frequency voltages: $U_{\text{rf}} = 125$ V (triangles), 145 V (circles) and 185 V (squares). We note that the vertical error bars on some points are smaller than the size of the data points.

temperature from ~ 7 K to ~ 60 K, and hence cover temperatures both considerably below and considerably above the steady-state temperature without the buffer gas (33 K). For the case of single molecular ions, the rotational temperature may be tuned as well, by momentarily moving the ion into regions with large, but controlled, micromotion.

An alternative method to reach low rotational temperatures is passive radiative cooling in a cryogenic environment. However, for the lowest temperature we report, such a scheme will require cooling periods several orders of magnitude longer, and it will furthermore be inapplicable for internal-state cooling of non-polar molecules, for which buffer-gas cooling is known to work²⁸.

We have also studied the cooling dynamics by measuring the times taken for the rotational ground state to refill after first being emptied by a set of REMP pulses. Because our REMP laser system is currently limited to a 10-Hz repetition rate, observation of the refilling dynamics required a decrease in the He buffer-gas pressure by a factor of ten relative to the values used for the experimental results presented in Figs 1–3. In Fig. 4, we plot the rotational ground-state population as a function of the delay of a second set of REMP pulses after full depletion of the ground state. From the fit to the data points, we find a

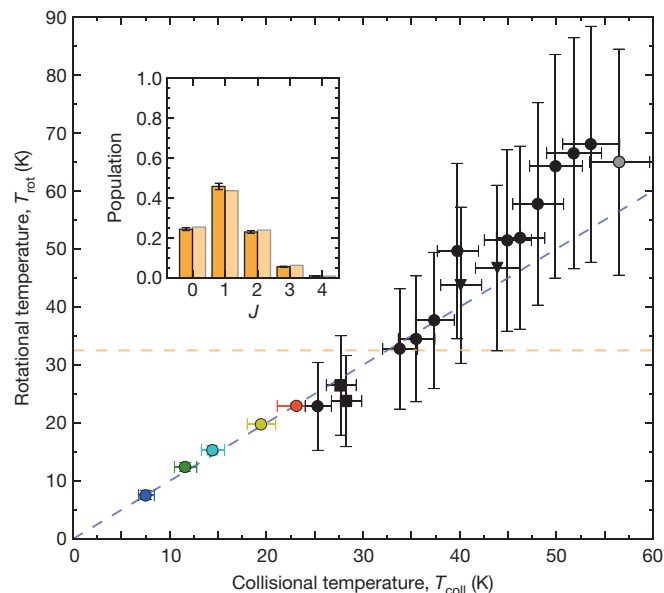


Figure 3 | Rotational temperature versus collisional temperature.

Rotational temperatures, T_{rot} , corresponding to the measured $J=0$ populations as in Fig. 2b, with the average micromotion energy replaced with the effective collisional temperature (see text and Supplementary Information). Error bars, 1 s.d. The light purple dashed line indicates $T_{\text{rot}} = T_{\text{coll}}$, and the horizontal dashed orange line indicates the rotational temperature in the case of no buffer gas ($T = 33$ K). Inset: the corresponding rotational-state distribution. Error bars, 1 s.e.m. We note that the vertical error bars on some points are smaller than the size of the data points.

refilling rate of about 1 s^{-1} , which for the above experimental condition translates into a rate of 10 s^{-1} . With a He temperature of 8.7 K, and assuming the cooling rate to be roughly equal to the Langevin collision rate, we estimate the He density and pressure to be 10^{10} cm^{-3} and 10^{-8} mbar, respectively, or four to five orders of magnitude lower than usually used in buffer-gas cooling experiments^{7,11}. Owing to this extremely low pressure, the He gas does not preclude sympathetic crystallization of the MgH^+ ions initially formed with kinetic energies in the range of ~ 0.1 eV. Moreover, the method should also allow for the introduction of molecular ions from an external source. We have indirectly tested

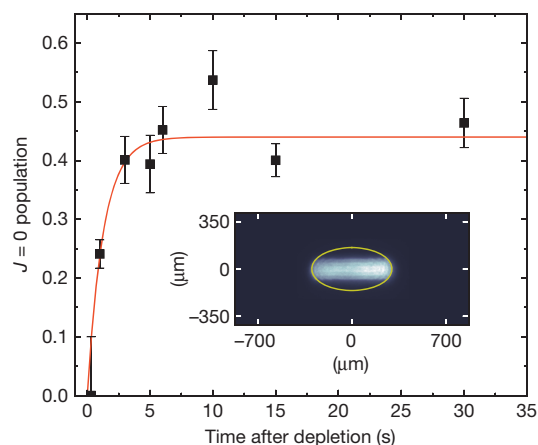


Figure 4 | Cooling dynamics. Measured rotational ground-state population at different delays after its depletion. Error bars, 1 s.d. The He density was roughly ten times lower than that used in producing the results presented in Figs 1–3, to be able to resolve the dynamics with the low repetition rate (10 Hz) of the REMP laser. The red curve is an exponential fit to the data, yielding a refilling rate of $0.8 \pm 0.2 \text{ s}^{-1}$. From this rate, we extrapolate the rate under the experimental conditions of Figs 1–3 to be $\sim 10 \text{ s}^{-1}$. Inset: projection image of the crystal with $(N_{\text{Mg}}, N_{\text{MgH}}) = (650, 950)$ used for this refilling data.

this possibility by producing a single, non-laser-cooled $^{25}\text{Mg}^+$ ion with a kinetic energy of several electronvolts by means of isotope-selective photoionization, and sympathetically crystallizing it with a single, laser-cooled $^{24}\text{Mg}^+$ ion in the presence of the He buffer gas. Finally, we have demonstrated that it is possible to increase the buffer-gas pressure, and thus the cooling rate, by a factor of at least five relative to the conditions leading to the results presented in Figs 1–3. Such higher cooling rates may be important for efficient rovibrational cooling of larger molecules with significantly more internal motional degrees of freedom.

The low buffer-gas collision rate achieved in the present experiment should enable investigations of laser-induced coherent processes as a function of the rotational temperature at the single-molecular level by momentarily moving the molecular ion away from the radio-frequency nodal line of the trapping potential. Applied to single molecules of biological interest, this could potentially be important for unravelling complex energy transport mechanisms such as those governing the coherent phenomena observed in light-harvesting molecules²⁹. Even lower temperatures may potentially be reached using a buffer gas consisting of laser-cooled atoms. Although challenges remain in relation to potential reactive collisions, recent results have demonstrated vibrational cooling of BaCl^+ ions from 300 K to 200 K by means of a gas of calcium atoms laser-cooled to a few millikelvin³⁰.

In the future, the technique presented here could even open up the possibility of combining internal-state buffer-gas cooling with translational sympathetic sideband cooling. Combined with a probabilistic state-preparation scheme for single-molecular ions³¹, this could bring within reach complete quantum-state preparation of a broad range of molecules for spectroscopic and ultracold ion-chemistry studies.

Received 23 October; accepted 19 December 2013.

Published online 9 March 2014.

- Hudson, E. R., Lewandowski, H. J., Sawyer, B. C. & Ye, J. Cold molecule spectroscopy for constraining the evolution of the fine structure constant. *Phys. Rev. Lett.* **96**, 143004 (2006).
- Hudson, J. J. *et al.* Improved measurement of the shape of the electron. *Nature* **473**, 493–496 (2011).
- Kamrath, M. Z., Relph, R. A., Guasco, T. L., Leavitt, C. M. & Johnson, M. A. Vibrational predissociation spectroscopy of the H_2 -tagged mono- and dicarboxylate anions of dodecanedioic acid. *Int. J. Mass Spectrom.* **300**, 91–98 (2011).
- Rizzo, T. R., Park, Y. D., Peteanu, L. A. & Levy, D. H. The electronic spectrum of the amino acid tryptophan in the gas phase. *J. Chem. Phys.* **84**, 2534–2541 (1986).
- Stearns, J. A. *et al.* Conformation-specific spectroscopy and photodissociation of cold, protonated tyrosine and phenylalanine. *J. Am. Chem. Soc.* **129**, 11814–11820 (2007).
- Willitsch, S., Bell, M. T., Gingell, A. D. & Softley, T. P. Chemical applications of laser- and sympathetically-cooled ions in ion traps. *Phys. Chem. Chem. Phys.* **10**, 7200–7210 (2008).
- Carr, L. D., DeMille, D., Krems, R. V. & Ye, J. Cold and ultracold molecules: science, technology and applications. *New J. Phys.* **11**, 055049 (2009).
- Smith, I. W. M. *Low Temperatures and Cold Molecules* (Imperial College Press, 2008).
- Hutzler, N. R. A cryogenic beam of refractory, chemically reactive molecules with expansion cooling. *Phys. Chem. Chem. Phys.* **13**, 18976–18985 (2011).
- Scoles, G., Bassi, D., Buck, U. & Laine, D. C. *Atomic and Molecular Beam Methods* Vol. 1 (Oxford Univ. Press, 1988).
- Gerlich, D. Ion-neutral collisions in a 22-pole trap at very low energies. *Phys. Scr.* **T59**, 256–263 (1995).
- Weinstein, J. D., deCarvalho, R., Guillet, T., Friedrich, B. & Doyle, J. M. Magnetic trapping of calcium monohydride molecules at millikelvin temperatures. *Nature* **395**, 148–150 (1998).
- Mølhave, K. & Drewsen, M. Formation of translationally cold MgH^+ and MgD^+ molecules in an ion trap. *Phys. Rev. A* **62**, 011401(R) (2000).
- Staanum, P. F., Højbjerg, K., Wester, R. & Drewsen, M. Probing isotope effects in chemical reactions using single ions. *Phys. Rev. Lett.* **100**, 243003 (2008).
- Højbjerg, K. *et al.* Consecutive photodissociation of a single complex molecular ion. *Phys. Rev. A* **77**, 030702(R) (2008).
- Zipkes, C., Palzer, S., Sias, C. & Köhl, M. A trapped single ion inside a Bose–Einstein condensate. *Nature* **464**, 388–391 (2010).
- Schmid, S., Härter, A. & Denschlag, J. H. Dynamics of a cold trapped ion in a Bose–Einstein condensate. *Phys. Rev. Lett.* **105**, 133202 (2010).
- Staanum, P. F., Højbjerg, K., Skyt, P. S., Hansen, A. K. & Drewsen, M. Rotational laser cooling of vibrationally and translationally cold molecular ions. *Nature Phys.* **6**, 271–274 (2010).
- Schneider, T., Roth, B., Duncker, H., Ernsting, I. & Schiller, S. All-optical preparation of molecular ions in the rovibrational ground state. *Nature Phys.* **6**, 275–278 (2010).
- Tong, X., Winney, A. H. & Willitsch, S. Sympathetic cooling of molecular ions in selected rotational and vibrational states produced by threshold photoionization. *Phys. Rev. Lett.* **105**, 143001 (2010).
- Ostendorf, A. *et al.* Sympathetic cooling of complex molecular ions to millikelvin temperatures. *Phys. Rev. Lett.* **97**, 243005 (2006).
- Hansen, A. K., Sørensen, M. A., Staantum, P. F. & Drewsen, M. Single-ion recycling reactions. *Angew. Chem. Int. Ed.* **51**, 7960–7962 (2012).
- Versolato, O. O. *et al.* Decay rate measurement of the first vibrationally excited state of MgH^+ in a cryogenic Paul trap. *Phys. Rev. Lett.* **111**, 053002 (2013).
- Blümel, R., Kappler, C., Quint, W. & Walther, H. Chaos and order of laser-cooled ions in a Paul trap. *Phys. Rev. A* **40**, 808–823 (1989).
- Schmidt, P. O. *et al.* Spectroscopy using quantum logic. *Science* **309**, 749–752 (2005).
- Hlavenka, P. *et al.* Absolute photodetachment cross section measurements of the O^- and OH^- anion. *J. Chem. Phys.* **130**, 061105 (2009).
- Schwarz, M. *et al.* Cryogenic linear Paul trap for cold highly charged ion experiments. *Rev. Sci. Instrum.* **83**, 083115 (2012).
- Schlemmer, S., Kuhn, T., Lescop, E. & Gerlich, D. Laser excited N_2^+ in a 22-pole ion trap: experimental studies of rotational relaxation processes. *Int. J. Mass Spectrom.* **185–187**, 589–602 (1999).
- Engel, G. S. *et al.* Evidence for wavelike energy transfer through quantum coherence in photosynthetic systems. *Nature* **446**, 782–786 (2007).
- Reilger, W. G. *et al.* Evidence for sympathetic vibrational cooling of translationally cold molecules. *Nature* **495**, 490–494 (2013).
- Vogelius, I. S., Madsen, L. B. & Drewsen, M. Probabilistic state preparation of a single molecular ion by projection measurement. *J. Phys. B* **39**, S1259–S1265 (2006).

Supplementary Information is available in the online version of the paper.

Acknowledgements M.D. appreciates generous support through the Danish National Research Foundation Center for Quantum Optics – QUANTOP; The Danish Agency for Science, Technology and Innovation; the Carlsberg Foundation; the Lundbeck Foundation; and the European Commission under the Seventh Framework Programme FP7 GA 607491 COMIQ. O.O.V., M.S., L.K. and A.W. acknowledge funding from STSM travel grants from COST-Action IOTA. Finally, the MPIK mechanical workshops have been of crucial importance for the construction of the cryogenic trap.

Author Contributions L.K., O.O.V., A.K.H., S.B.K. and A.G. contributed equally to the work by carrying out the experiments, taking part in the data analysis and writing the manuscript. M.S., A.W., J.U. and J.R.C.L.-U. contributed significantly by designing and constructing the cryogenically cooled trap. M.D. contributed to the trap design, had the idea for the experiment, devised the experimental protocol, led the data analysis and wrote most of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.D. (drewsen@phys.au.dk).

Coherent control of the waveforms of recoilless γ -ray photons

Farit Vagizov^{1,2,3}, Vladimir Antonov^{1,4,5}, Y. V. Radeonychev^{1,4,5}, R. N. Shakhmurov^{2,3} & Olga Kocharovskaya¹

The concepts and ideas of coherent, nonlinear and quantum optics have been extended to photon energies in the range of 10–100 kilo-electronvolts, corresponding to soft γ -ray radiation (the term used when the radiation is produced in nuclear transitions) or, equivalently, hard X-ray radiation (the term used when the radiation is produced by electron motion). The recent experimental achievements in this energy range include the demonstration of parametric down-conversion in the Langevin regime¹, electromagnetically induced transparency in a cavity², the collective Lamb shift³, vacuum-assisted generation of atomic coherences⁴ and single-photon revival in nuclear absorbing multilayer structures⁵. Also, realization of single-photon coherent storage⁶ and stimulated Raman adiabatic passage⁷ were recently proposed in this regime. More related work is discussed in a recent review⁸. However, the number of tools for the coherent manipulation of interactions between γ -ray photons and nuclear ensembles remains limited. Here we suggest and implement an efficient method to control the waveforms of γ -ray photons coherently. In particular, we demonstrate the conversion of individual recoilless γ -ray photons into a coherent, ultrashort pulse train and into a double pulse. Our method is based on the resonant interaction of γ -ray photons with an ensemble of nuclei with a resonant transition frequency that is periodically modulated in time. The frequency modulation, which is achieved by a uniform vibration of the resonant absorber, owing to the Doppler effect, renders resonant absorption and dispersion both time dependent, allowing us to shape the waveforms of the incident γ -ray photons. We expect that this technique will lead to advances in the emerging fields of coherent and quantum γ -ray photon optics, providing a basis for the realization of γ -ray-photon/nuclear-ensemble interfaces and quantum interference effects at nuclear γ -ray transitions.

Quantum optics deals with the interaction of photons with the quantum transitions of matter, providing the basis for the new, fast-growing fields of quantum cryptography, quantum communication and quantum information (see, for example, ref. 9 and references therein). So far, experiments in these fields have been implemented with either microwave or optical photons interacting with atomic transitions, and typically require cryogenic temperatures.

The γ -ray photons (for which we use the shorthand ' γ -photons') in the 10–100-keV energy range have important potential advantages over the microwave and optical photons for applications in quantum cryptography, communication and information, for a number of reasons. First, with such photons there is nearly 100% detector efficiency and low noise (almost no false detections). Second, there is extremely high (potentially sub-ångström and, at the moment, nanometre¹⁰) spatial resolution. This removes what is now a fundamental limitation on the creation of nanoscale photonic circuits set by the diffraction limit, which is $\sim 1 \mu\text{m}$ for optical photons. Third, they have a high penetration depth in many materials that are opaque to optical photons. Fourth, with such photons the capacity of information channels is potentially much higher.

Nuclear transitions in solids in this energy range have a number of advantages compared with electronic optical transitions. They may present perfect two-level systems with very narrow linewidths (for example 1.1 MHz for the 14.4-keV transition in ⁵⁷Fe) owing to the Mössbauer effect (even at room temperature) and deep shielding from the environment. The quality factor for the 14.4-keV transition is $Q = 3 \times 10^{12}$, as per definition: Q is the ratio of resonance energy to bandwidth. Coherent multiple scattering of a γ -photon in an optically thick, resonant absorber enables potential applications for single γ -photon/nuclear-ensemble interfaces owing to the collective enhancement of the photon–nucleus interaction. We note that in the optical range such single-photon/atomic-ensemble interfaces have been developed recently as a powerful alternative to cavity-enhanced single-photon/single-atom interfaces¹¹. Moreover, natural sources of single γ -photons exist in the form of low-activity radioactive sources, for which discrete emissions well separated in time (compared with the natural lifetime of the level) ensure the single-photon nature of the radiation. Also, the cascade scheme of the decay, characteristic of some radioactive Mössbauer sources (for example ⁵⁷Co; Fig. 1a), ensures heralding of the emission of the second γ -photon by detection of the first γ -photon in the cascade, which 'starts the clock' for measuring the waveform of the second photon¹².

The methods for coherently controlling temporal waveforms of single photons (well developed in optics owing to the use of high-finesse cavities and bright, coherent sources of radiation) are limited in γ -optics, although the emerging state-of-the-art facilities, such as self-seeded hard X-ray free-electron lasers¹³ and nearly 100% reflecting mirrors¹⁴, will provide an impetus for the development of such techniques.

We propose an efficient method to shape γ -photons in an optically thick, uniformly vibrating, resonant, recoilless absorber, and perform a

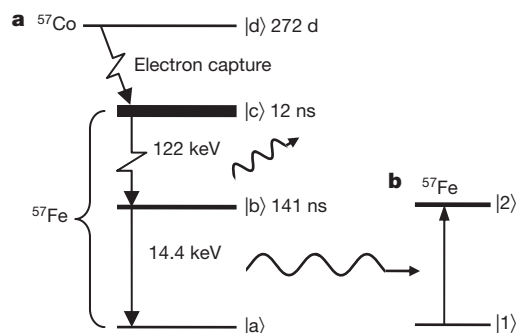


Figure 1 | Energy scheme of emission and resonant absorption of a 14.4-keV photon. **a**, The ⁵⁷Co nuclide in the state $|d\rangle$ decays with half-life of $T_{1/2} = 272$ d, producing an ⁵⁷Fe nucleus in the state $|c\rangle$, which decays with a lifetime of $T_c = 12$ ns to the state $|b\rangle$, emitting a 122-keV γ -photon. The state $|b\rangle$ decays with lifetime of $T_b = 141$ ns to the ground state, $|a\rangle$, emitting a 14.4-keV photon. **b**, The 14.4-keV transition between the ground state, $|1\rangle$, and the first excited state, $|2\rangle$, of the absorbing ⁵⁷Fe nucleus with the linewidth 1.13 MHz.

¹Department of Physics and Astronomy and Institute for Quantum Science and Engineering, Texas A&M University, College Station, Texas 77843-4242, USA. ²Kazan Federal University, Kazan 420008, Russia. ³Kazan Physical-Technical Institute of the Russian Academy of Sciences, Kazan 420029, Russia. ⁴Institute of Applied Physics of the Russian Academy of Sciences, Nizhny Novgorod 603950, Russia. ⁵N.I. Lobachevsky State University, Nizhny Novgorod 603950, Russia.

proof-of-principle experiment with a ^{57}Co source of single γ -photons and a ^{57}Fe absorber (Fig. 1).

The basic idea is that the vibration of an absorber leads to periodic modulation of the resonant $|1\rangle \leftrightarrow |2\rangle$ nuclear transition frequency (Fig. 1b) with respect to the frequency of the incident photons owing to the Doppler effect. As a result, the quasi-monochromatic incident radiation is transformed during its propagation into a spectral comb at the point where it exits the absorber. The relative amplitudes and phases of the produced spectral components are defined by the vibration amplitude and frequency, the detuning of the central frequency of the source from the resonant frequency of the absorber, the source and absorber linewidths, and the absorber optical depth. By changing these parameters, we can adjust the amplitudes and phases of the output spectral components and produce waveforms different from the incident γ -photons. In particular, the creation of a constant phase difference between equidistant neighbouring spectral components of comparable amplitude would result in a sequence of pronounced, bandwidth-limited pulses.

The discussed method to produce ultrashort γ -ray pulses essentially constitutes a γ -ray implementation of a general approach based on the ultrafast variation of the parameters of a resonant quantum transition (see ref. 15 and references therein). Whereas in the case of atomic transitions the modulation of the transition frequency was produced by the quasi-static Stark effect in a strong laser field, in the case of nuclear transitions it is due to the Doppler effect. This resonant parametric technique, unlike the traditional techniques of nonlinear optics, does not require a high intensity of incoming radiation and, hence, can be implemented even in the single-photon regime.

The frequency modulation of recoilless emission by mechanical vibration has been intensively studied since the late 1970s in applications to spectroscopic measurements and the calibration of mechanical displacements (see refs 16, 17 and references therein). Recently, there was a suggestion to produce ultrashort γ -pulses using a vibrating source, emitting frequency-modulated radiation, with a far-off-resonance absorber compensating the phase mismatch of the incident γ -radiation spectral components¹⁸. Although sound in principle, that proposal is difficult to implement experimentally, because weak non-resonant interaction requires an absorber so thick that off-resonance losses are undesirably large. Use of a resonant absorber resolves that problem. Also, in practice it is easier to realize uniform oscillations of a thin, resonant absorber than of a radiative source.

The scheme of our experimental set-up is given in Fig. 2. Observation of a single-photon waveform, shaped by the vibrating, resonant absorber, is based on the time-delayed coincidence counts of the sequentially emitted 122-keV and 14.4-keV γ -photons (Fig. 1a). Registering a 122-keV photon defines the initial moment, t_0 , of the formation of the 14.4-keV state and the beginning of photon emission (equation (2); all numbered equations are in Methods). We measure the delay, $\tau = t - t_0$, in the detection of the 14.4-keV photon at the exit of the absorber with a 5-ns bin width, repeating this procedure many times. As a result, we obtain the coincidence count rate, as a function of delay, $N(\tau)$, corresponding to the single-photon waveform $|\Psi(\tau)|^2$ (refs 5, 12, 16–21). These experiments are essentially inspired by the recent demonstration of a new technique for optical single-photon shaping²².

Transmission through the absorber vibrating with frequency Ω results in the reshaping of the exponential waveform of the incident photon. The output shape depends on the vibration phase, $\vartheta_0 = \Omega t_0$, at time t_0 , but because emission of 122-keV photon is a stochastic process, ϑ_0 is random. To reveal the coherent effect of the absorber vibration on the 14.4-keV-photon waveform, we collect only the coincidence counts of 122-keV photons and those 14.4-keV photons whose 122-keV photons were detected within short intervals, dt , around the times $t_0^{(n)} = (\vartheta_0 + 2\pi n)/\Omega$, where $n = 0, 1, 2, \dots$ (Fig. 2), corresponding to the same vibration phase, ϑ_0 .

This technique allowed us to demonstrate the temporal compression of an individual γ -photon into a decaying train of ultrashort pulses that were an order of magnitude shorter than the decay time of the

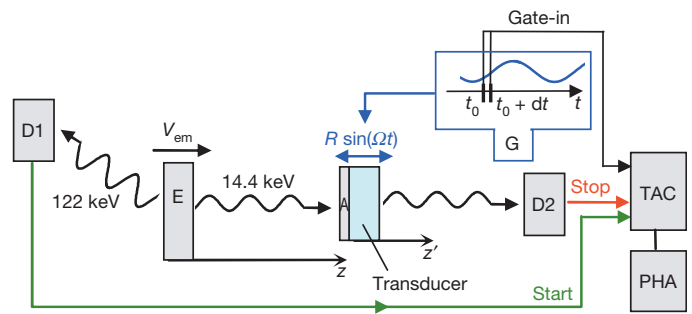


Figure 2 | Experimental set-up for γ -photon waveform control. The emitter (E) is a radioactive $^{57}\text{Co}:\text{Rh}$ foil. It can be moved by the Mössbauer transducer with tunable constant velocity, V_{em} , relative to the absorber (A), which is 25- μm -thick stainless-steel foil with a natural abundance ($\sim 2\%$) of ^{57}Fe (optical depth, $T_M = 5.18$). It is glued on the polyvinylidene fluoride piezo-transducer that transforms the sinusoidal signal from the radio-frequency generator (G) into the uniform vibration (R , vibration amplitude). The time-to-amplitude converter (TAC), operating in the coincidence mode, receives the 'start' pulse, produced by detector D1 on registration of a 122-keV photon, and the 'stop' pulse, produced by detector D2 on registration of a 14.4-keV photon transmitted by the vibrating absorber. The delay, τ , between the stop and start pulses is measured by the pulse-height analyser (PHA). At times $t_0^{(n)}$, matching the chosen vibration phase, $\Omega t_0^{(n)} = \vartheta_0 + 2\pi n$, the generator also produces 'gate-in' signals, allowing TAC and PHA to measure the τ provided that the 122-keV photon is received only within the short interval between $t_0^{(n)}$ and $t_0^{(n)} + dt$.

excited nuclear state (Fig. 3b). In the following, we show experimental results that were obtained under conditions leading to pulses with maximum peak height for the given absorber thickness.

The physical origin of the waveform transformation is revealed by spectral analysis. The spectrum (Fig. 3a, inset) of the quasi-monochromatic incoming photon (Fig. 3b, red dashed curve) in the laboratory reference frame is 'seen' by the vibrating nuclei of the absorber as a comb of equidistant spectral components separated by the vibration frequency, Ω (Fig. 3a and equations (4) and (5)). Under the chosen modulation index, $p = 1.8$ (see Methods), there are seven major spectral components. It can be seen that the upper-frequency sidebands are phase-matched with a phase difference of π between each neighbouring pair (brown squares), whereas three lower-frequency sidebands are phase matched with a phase difference of 0 (green triangles). Taking into account the equivalence of $2\pi n$ -shifted phases of the -2 and $+2$ sidebands (Fig. 3a), we can match five major spectral components, eliminating either the -1 or the $+1$ sideband by tuning it to the absorber resonance (by properly choosing the constant velocity of the source with respect to the absorber). This leads to the waveforms shown by the red line in Fig. 3b (elimination of -1) and blue line in the inset of Fig. 3b (elimination of $+1$). In terms of the shift of the produced pulse train relative to the front edge of the photon, these waveforms differ from each other by half the vibration cycle, π/Ω . Owing to constructive interference of the generated sidebands, the peak values of the count rate rise above the exponential waveform of the incident photon. Even higher peak heights will be achieved if, instead of suppression of the phase-mismatched sideband through resonant absorption, its phase is shifted by π through resonant dispersion. Such a phase shift requires detuning from the exact resonance and a several-fold greater resonant optical depth; without strong off-resonance photoelectric absorption, it can be realized in a stainless-steel film highly enriched with ^{57}Fe .

We also demonstrate the possibility of a single photon splitting into two pulses (Fig. 3c), for the first time realizing a time-bin qubit in the γ -ray frequency range²³. The time interval between the pulses and the ratio of their amplitudes can be controlled by the vibration phase (Methods).

A direct count rate of the whole flow of the 14.4-keV photons versus time starting from $t_0^{(n)}$ (without binding to the 122-keV photons) corresponds to averaging the 14.4-keV-photon count rate, $N(t - t_0)$, over the time t_0 (equation (12)). In this case, the random flow of γ -photons

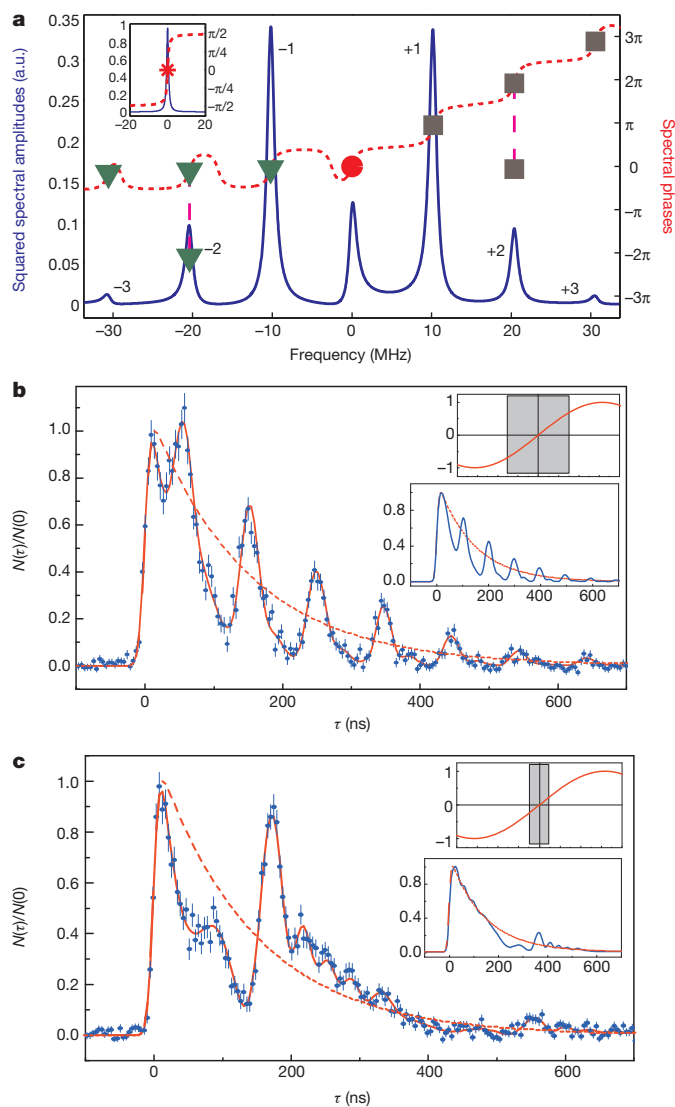


Figure 3 | Shaping of γ -photon. **a**, Spectrum of the incident 14.4-keV-photon waveform in the reference frame of the vibrating absorber and in the laboratory reference frame (inset), calculated according to equation (5) (blue solid and red dashed lines are spectral amplitudes and phases, respectively). Central phases of the lower-frequency, principle-frequency and upper-frequency components are labelled by triangles, circle and squares, respectively. The vibration frequency, phase and amplitude (modulation index) are respectively $\Omega/2\pi = 10.2$ MHz, $\vartheta_0 = 0$ and $R = 0.25$ Å ($p = 1.8$). a.u., arbitrary units. **b**, The normalized coincidence count rate of 14.4-keV photons, $N(\tau)/N(0)$, passed through the vibrating absorber (γ -photon waveform). The ‘−1’ sideband is tuned to the absorber spectral line by moving the emitter towards the absorber with velocity $V_{\text{em}} = 0.88$ mm s $^{-1}$. The experimental data are blue dots, the background is subtracted and the error bars are estimated from photon counting statistics as $\sqrt{N(\tau)/N(0)}$. Red solid and dashed curves are plotted according to equation (11) and equation (3), respectively. Upper inset, sinusoidal vibration of the absorber and the acquisition interval, $\Omega\Delta t = \pi/2$, limited by the equipment used. Lower inset, the single-photon waveform in the case where the ‘+1’ sideband is tuned to the absorber resonance. **c**, The 14.4-keV-photon waveform in the case where $\Omega/2\pi = 2.6$ MHz, $\vartheta_0 = -0.3\pi$, $R = 0.9$ Å ($p = 6.58$), the emitter is moved towards the absorber with velocity $V_{\text{em}} = 1.13$ mm s $^{-1}$ (tuning the ‘−5’ sideband to the absorber resonance), and the acquisition interval is $\pi/8$ (upper inset). Lower inset, the 14.4-keV-photon waveform under the same conditions, except that $V_{\text{em}} = -1.13$ mm s $^{-1}$ (tuning the ‘+5’ sideband to the absorber resonance).

transforms into a train of nanosecond pulses (Fig. 4). The pulse repetition period is equal to $2\pi/\Omega$, and the pulse duration is defined by the inverse of the product of the number of phase-matched components

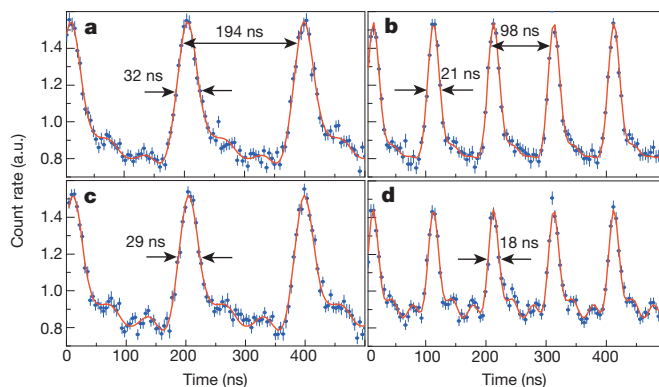


Figure 4 | Pulses of Mössbauer radiation. Count rate of the 14.4-keV photons versus time in the case where detector D1 is removed and the ‘gate-in’ signal, matched with the zero phase of vibration, triggers TAC and PHA to measure time elapsed until detector D2 produces a ‘stop’ signal on registration of 14.4-keV photon. Repeating this procedure many times results in a count rate of the 14.4-keV photons versus time corresponding to the intensity of 14.4-keV radiation. The constant velocity, V_{em} , of the emitter towards the absorber tunes the ‘−1’ sideband to the absorber resonance (Fig. 3a). **a**, $\Omega/2\pi = 5.16$ MHz, $V_{\text{em}} = 0.44$ mm s $^{-1}$, $R = 0.25$ Å, ($p = 1.8$); **b**, $\Omega/2\pi = 10.2$ MHz, $V_{\text{em}} = 0.87$ mm s $^{-1}$, $R = 0.25$ Å ($p = 1.8$); **c**, $\Omega/2\pi = 5.16$ MHz, $V_{\text{em}} = 0.44$ mm s $^{-1}$, $R = 0.27$ Å ($p = 2$); **d**, $\Omega/2\pi = 10.2$ MHz, $V_{\text{em}} = 0.87$ mm s $^{-1}$, $R = 0.27$ Å ($p = 2$). The experimental data are blue dots and the error bars are estimated from photon counting statistics as above. The red solid line is plotted according to equations (13) and (14).

(proportional to the modulation index) and the modulation frequency. An increase in the vibration frequency at a fixed vibration amplitude results in a proportional shortening of both the pulse duration and the repetition period, owing to an increase in the separation between the sidebands (compare Fig. 4a with Fig. 4b and Fig. 4c with Fig. 4d). An increase in the vibration amplitude at a fixed vibration frequency results in a pulse shortening without change in the repetition period, owing to an increase in the number of sidebands (compare Fig. 4a with Fig. 4c and Fig. 4b with Fig. 4d). However, the short, intense pulses are accompanied by the additional weak pulses appearing as a result of the phase-mismatched components.

The maximum potentially available vibration frequency is 1 GHz with rather small modulation index, $p < 2$. A much higher modulation index ($p > 10$) may be achieved with moderate modulation frequencies (10–100 MHz), but in this case additional care should be taken in phase-matching the produced sidebands (using, for example, a chain of properly tuned resonantly absorbing foils or external phase synchronization by the diffraction gratings). Thus, the shortest pulses, which potentially could be produced by that technique, are limited to 100 ps (unless other methods providing higher modulation frequency and higher modulation index of the nuclear transition are used).

The described set-up is at present the only known tabletop source of ultrashort γ -ray pulses. The parameters of the pulses, including the shape and number of pulses, the repetition rate (potentially variable in the megahertz–gigahertz range) and the duration (potentially 100 ns–100 ps), can be widely controlled. This allows for a variety of time-resolved experiments (including dynamic X-ray diffraction) directly in local labs, although such a tabletop set-up cannot compete, in terms of pulse duration and photon flux, with modern synchrotrons²⁴. However, it has two unique features leading to applications that are unavailable with other existing sources of ultrashort X-ray pulses.

First, the produced pulses have exclusive spectral–temporal properties. At the moment, they are the only nearly transform-limited ultrashort pulses that can be produced in this frequency range. They can be used to generate frequency combs with a fixed frequency interval and phase-matched spectral components, providing a unique combination of wide spectral coverage and narrow spectral width for each component.

Such pulses are promising for the realization of frequency comb spectroscopy (broadband, high-spectral-resolution spectroscopy) of nuclei in solids in the 10–100-keV energy range as well as time-resolved Mössbauer spectroscopy. They can also be used for efficient preparation of a coherent superposition of nuclear states, and for realization of quantum interference effects such as modulation-induced transparency²⁵ and electromagnetically induced transparency^{26,27}. The last effect has numerous applications in the optical frequency range, such as the implementation of resonantly enhanced nonlinearities (including single-photon nonlinearities), slow and stored light, lasing without inversion, and so on (see ref. 28 and references therein). Similar applications may be anticipated in the γ -ray range.

Second, at the moment our tabletop set-up is the only source of single-photon, ultrashort pulses with efficiently controllable waveforms. The splitting of the single photon into two pulses constitutes the first realization of a time-bin qubit in this range of frequencies. The production of single-photon, ultrashort-pulse trains with pulse duration much shorter than the natural lifetime of the emitting nuclear level and with controllable waveforms provides a unique opportunity for the realization of quantum memories^{29,30} and other nuclear-ensemble/ γ -photon interfaces, opening the way to applications in quantum communication and information.

METHODS SUMMARY

The transformation of a recoilless, spontaneously emitted γ -photon on its propagation through a uniformly vibrating, resonant Mössbauer absorber is described semiclassically by numerically and analytically solving the Maxwell–Bloch equations with a harmonically modulated nuclear transition frequency in the rotating-wave and linear approximations. Modification of the shape of the γ -photon at the exit of the absorber is studied as a function of the parameters of the system, such as the frequency and amplitude of modulation, the detuning of the central frequency of the source from the resonance frequency of the absorber, and the vibration phase at the time of the γ -photon arriving at the absorber entrance.

We numerically plot waveforms for different parameter values to demonstrate how to produce a variety of waveforms from a single γ -photon (including a decaying train of ultrashort pulses, a double pulse and a single triangular pulse). The regime of formation of a train of γ -ray pulses with pulse-to-pulse coherence and a duration an order of magnitude less than the natural relaxation time of the resonant nuclear transition is described analytically within a simplified model.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 25 August 2013; accepted 13 January 2014.

Published online 16 March 2014.

1. Schwartz, S. *et al.* X-ray parametric down-conversion in the Langevin regime. *Phys. Rev. Lett.* **109**, 013602 (2012).
2. Röhlberger, R., Wille, H.-C., Schlage, K. & Sahoo, B. Electromagnetically induced transparency with resonant nuclei in a cavity. *Nature* **482**, 199–203 (2012).
3. Röhlberger, R., Schlage, K., Sahoo, B., Couet, S. & Rüffer, R. Collective Lamb shift in single-photon superradiance. *Science* **328**, 1248–1251 (2010).
4. Heeg, K. P. *et al.* Vacuum-assisted generation and control of atomic coherences at x-ray energies. *Phys. Rev. Lett.* **111**, 073601 (2013).
5. Shakhmuratov, R., Vagizov, F. & Kocharovskaya, O. Single γ -photon revival and radiation burst in a sandwich absorber. *Phys. Rev. A* **87**, 013807 (2013).

6. Liao, W.-T., Pálffy, A. & Keitel, C. H. Coherent storage and phase modulation of single hard-X-ray photons using nuclear excitons. *Phys. Rev. Lett.* **109**, 197403 (2012).
7. Liao, W.-T., Pálffy, A. & Keitel, C. H. Nuclear coherent population transfer with X-ray laser pulses. *Phys. Lett. B* **705**, 134–138 (2011).
8. Adams, B. W. *et al.* X-ray quantum optics. *J. Mod. Opt.* **60**, 2–21 (2013).
9. Ma, X.-S. *et al.* Quantum teleportation over 143 kilometres using active feed-forward. *Nature* **489**, 269–273 (2012).
10. Döring, F. *et al.* Sub-5 nm hard X-ray point focusing by a combined Kirkpatrick-Baez mirror and multilayer zone plate. *Opt. Express* **21**, 19311–19323 (2013).
11. Hammerer, K., Sørensen, A. S. & Polzik, E. S. Quantum interface between light and atomic ensembles. *Rev. Mod. Phys.* **82**, 1041–1093 (2010).
12. Hoy, G. R., Hamill, D. W. & Wintersteiner, P. P. in *Mössbauer Effect Methodology* Vol. 6 (ed. Gruverman, I. J.) 109–121 (Plenum, 1970).
13. Amann, J. *et al.* Demonstration of self-seeding in a hard-X-ray free-electron laser. *Nature Photon.* **6**, 693–698 (2012).
14. Shvyd'ko, Yu., Stoupin, S., Blank, V. & Terentyev, S. Near-100% Bragg reflectivity of X-rays. *Nature Photon.* **5**, 539–542 (2011).
15. Antonov, V. A., Radeonychev, Y. V. & Kocharovskaya, O. Formation of a single attosecond pulse via interaction of resonant radiation with a strongly perturbed atomic transition. *Phys. Rev. Lett.* **110**, 213903 (2013).
16. Heliö, P., Titttonen, I. & Katila, T. Enhanced transient effects due to saturated absorption of recoilless γ -radiation. *Phys. Rev. B* **34**, 3458–3461 (1986).
17. Smirnov, G. V. & Potzel, W. Perturbation of nuclear excitons by ultrasound. *Hyperfine Interact.* **123/124**, 633–663 (1999).
18. Kuznetsova, E., Kolesov, R. & Kocharovskaya, O. Compression of γ -ray photons into ultrashort pulses. *Phys. Rev. A* **68**, 043825 (2003).
19. Smirnov, G. V. General properties of nuclear resonant scattering. *Hyperfine Interact.* **123/124**, 31–77 (1999).
20. Kagan, Yu. Theory of coherent phenomena and fundamentals in nuclear resonant scattering. *Hyperfine Interact.* **123/124**, 83–126 (1999).
21. Hannon, J. P. & Trammell, G. T. Coherent gamma-ray optics. *Hyperfine Interact.* **123/124**, 127–274 (1999).
22. Kolchin, P., Belthangady, C., Du, S., Yin, G. Y. & Harris, S. E. Electro-optic modulation of single photons. *Phys. Rev. Lett.* **101**, 103601 (2008).
23. Pittman, T. It's a good time for time-bin qubits. *Physics* **6**, 110 (2013).
24. Röhlberger, R. *Nuclear Condensed Matter Physics using Synchrotron Radiation* (Springer Tracts Mod. Phys. 208, Springer, 2005).
25. Radeonychev, Y. V., Tokman, M. D., Litvak, A. G. & Kocharovskaya, O. Acoustically induced transparency in optically dense resonance medium. *Phys. Rev. Lett.* **96**, 093602 (2006).
26. Kocharovskaya, O. A. & Khanin, Ya. I. Population trapping and coherent bleaching of a three-level medium by a periodic train of ultrashort pulses. *Sov. Phys. JETP* **63**, 945–949 (1986).
27. Harris, S. E. Electromagnetically induced transparency. *Phys. Today* **50**, 36–42 (1997).
28. Fleischhauer, M., Imamoglu, A. & Marangos, J. P. Electromagnetically induced transparency: optics in coherent media. *Rev. Mod. Phys.* **77**, 633–673 (2005).
29. Lvovsky, A. I., Sanders, B. C. & Tittel, W. Optical quantum memory. *Nature Photon.* **3**, 706–714 (2009).
30. Simon, C. *et al.* Quantum memories. *Eur. Phys. J. D* **58**, 1–22 (2010).

Acknowledgements We acknowledge the support from the US NSF (grant no. PHY-1307346), the RFBR (grants nos 13-02-00831 and 12-02-00263) and The Ministry of Education and Science of the Russian Federation (contract no. 11.G34.31.0011). V.A. acknowledges support from the Dynasty Foundation.

Author Contributions F.V. developed the experimental methods, designed the experiment and derived all the experimental results. V.A., Y.V.R. and R.N.S. developed the theoretical description. V.A. and F.V. determined the optimal parameters for the experiments and provided the theoretical fit to experimental data. Y.V.R. and F.V. suggested the technique for observing the single-photon waveforms. R.N.S. obtained analytical solutions for some limiting cases. O.K. suggested the idea, coordinated the efforts and wrote the paper. All authors discussed the results and edited the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to O.K. (kochar@physics.tamu.edu).

Highly siderophile elements in Earth's mantle as a clock for the Moon-forming impact

Seth A. Jacobson^{1,2}, Alessandro Morbidelli¹, Sean N. Raymond^{3,4}, David P. O'Brien⁵, Kevin J. Walsh⁶ & David C. Rubie²

According to the generally accepted scenario, the last giant impact on Earth formed the Moon and initiated the final phase of core formation by melting Earth's mantle. A key goal of geochemistry is to date this event, but different ages have been proposed. Some^{1–3} argue for an early Moon-forming event, approximately 30 million years (Myr) after the condensation of the first solids in the Solar System, whereas others^{4–6} claim a date later than 50 Myr (and possibly as late as around 100 Myr) after condensation. Here we show that a Moon-forming event at 40 Myr after condensation, or earlier, is ruled out at a 99.9 per cent confidence level. We use a large number of *N*-body simulations to demonstrate a relationship between the time of the last giant impact on an Earth-like planet and the amount of mass subsequently added during the era known as Late Accretion. As the last giant impact is delayed, the late-accreted mass decreases in a predictable fashion. This relationship exists within both the classical scenario^{7,8} and the Grand Tack scenario^{9,10} of terrestrial planet formation, and holds across a wide range of disk conditions. The concentration of highly siderophile elements (HSEs) in Earth's mantle constrains the mass of chondritic material added to Earth during Late Accretion^{11,12}. Using HSE abundance measurements^{13,14}, we determine a Moon-formation age of 95 ± 32 Myr after condensation. The possibility exists that some late projectiles were differentiated and left an incomplete HSE record in Earth's mantle. Even in this case, various isotopic constraints strongly suggest that the late-accreted mass did not exceed 1 per cent of Earth's mass, and so the HSE clock still robustly limits the timing of the Moon-forming event to significantly later than 40 Myr after condensation.

The Moon-forming impact must be the last giant impact experienced by Earth, because both Earth and Moon share an almost identical isotopic composition. This giant impact melted Earth's mantle and formed the final global magma ocean, causing core–mantle differentiation, and it possibly removed a significant portion of Earth's atmosphere. These events can be dated using radiometric chronometers. Unfortunately, the age of the Moon differs substantially from one chronometer to the next owing to assumptions in the computation of the so-called model ages. For instance, Touboul *et al.*⁴ measured very similar ¹⁸²W/¹⁸⁴W ratios for both Earth and the Moon and, given the differences in Hf/W ratios estimated at the time for the two bodies, concluded that the Moon-forming event must have been 62^{+90}_{-10} Myr after condensation. (Throughout, we use 'after condensation' to mean 'since the birth of the Solar System'; see online-only Methods for details.) In this way, the radioactive ¹⁸²Hf would have almost fully decayed into ¹⁸²W beforehand, thus easily accounting for the almost nonexistent difference in ¹⁸²W/¹⁸⁴W ratios between the Moon and Earth. However, König *et al.*¹⁵ subsequently determined that the Hf/W ratios in Earth and Moon are instead identical, voiding this reasoning and leaving the problem of dating the Moon-forming event wide open.

We approach this problem from a new direction, using a large number of *N*-body simulations of the accretion of the terrestrial planets from a disk of planetesimals and planetary embryos. The simulations extend

across the range of well-studied and successful scenarios (for example, typically creating the right number of planets on the correct orbits) and fall into two categories. First, classical simulations assume that the disk of terrestrial building blocks extended from an inner edge at 0.3–0.7 AU (where 1 AU is the Earth–Sun distance) from the Sun out to 4–4.5 AU, just interior to Jupiter, and that the giant planets stayed on orbits near their current ones. Our sample of 48 classical simulations produces 87 Earth-like planets^{7,8}, which are broadly defined as final bodies with masses within a factor of two of Earth's mass and with orbits between the current orbits of Mercury and Mars. In these simulations, the last giant impacts occur between 10 Myr and 150 Myr after the removal of the solar nebular gas, which happened about 3 Myr after condensation. The classical scenario has known shortfalls. The Mars analogues are too massive unless the giant planets are assumed to be initially on very eccentric orbits^{8,16}. These large eccentricities, however, cannot be explained in the context of the formation and evolution of giant planets in a gas disk. Furthermore, terrestrial planets accreted in the presence of very eccentric giant planets are very low in water^{7,17}.

The second category consists of simulations in which the terrestrial planets form from a disk with an abrupt outer edge at about 1 AU; the inner edge remains the same as in the classical simulations. Simulated solar systems in this second category successfully include Mars-like planets^{9,10,18}. The truncation of the outer edge can be explained by the early gas-driven inward-then-outward migration of Jupiter and Saturn known as the Grand Tack, which then produces a region of greatly depleted surface density between the current orbits of Mars and Jupiter⁹. These simulations best reproduce the orbital and mass distributions of the terrestrial planets, and they also explain the compositional structure of the asteroid belt⁹. Moreover, Earth-like planets accrete volumes of water that are consistent with estimates of Earth's water content¹⁰. Previously reported Grand Tack simulations feature last giant impacts occurring typically within about 50 Myr after the time of removal of the solar nebular gas^{9,10,18}. We complement those simulations with new ones (resulting in a total of 211 Grand Tack simulations, producing 354 Earth-like planets).

All Grand Tack simulations produce planetary systems that match the Solar System as well as or better than those obtained from classical simulations despite varying many initial disk conditions (see Extended Data Fig. 1), among them the initial total mass ratio between the embryo and planetesimal populations (from 1:1 to 8:1; see the online-only Methods). If the initial ratio of embryo mass to planetesimal mass is increased, the time of the last giant impact also increases (even to about 150 Myr after condensation, see Extended Data Fig. 2) owing to the reduction of the well-known effect of dynamical friction⁷—the damping of the eccentricities and inclinations of the larger bodies due to gravitational interactions with a swarm of smaller bodies. Higher eccentricities and inclinations of the embryos diminish mutual gravitational focusing, increasing the accretion timescale, and consequently leading to later embryo–embryo collisions (that is, giant impacts).

¹Observatoire de la Côte d'Azur, Laboratoire Lagrange, Boulevard de l'Observatoire, BP 4229, 06304 Nice Cedex 4, France. ²Universität Bayreuth, Bayerisches Geoinstitut, 95440 Bayreuth, Germany.

³Université Bordeaux, Laboratoire d'Astrophysique de Bordeaux, UMR 5804, 33270 Floirac, France. ⁴CNRS, Laboratoire d'Astrophysique de Bordeaux, UMR 5804, 33270 Floirac, France. ⁵Planetary Science Institute, 1700 East Fort Lowell, Suite 106, Tucson, Arizona 85719, USA. ⁶Southwest Research Institute, Planetary Science Directorate, 1050 Walnut Street, Suite 300, Boulder, Colorado 80302, USA.

Dynamical models alone do not indicate whether the Moon-forming impact occurred early (about 30 Myr after condensation) or late (about 50–100 Myr after condensation), because the result depends on the initial disk conditions. However, we find a clear statistical correlation between the time of the Moon-forming impact and the mass subsequently accreted, known as the late-accreted mass. This era of Late Accretion includes no giant impacts by definition and so all of the late-accreted mass comes from the planetesimal population. As shown in Fig. 1, this correlation exists across all simulations of both types: classical and Grand Tack. We interpret this correlation by considering that the planetesimal population decays over a characteristic time, so that if the last giant impact occurs earlier, then the remaining planetesimal population is larger. A larger remaining planetesimal population delivers a larger late-accreted mass. Strengthening the correlation, a larger initial planetesimal population leads to a shorter timescale for giant impacts owing to enhanced dynamical friction. For any given last-giant-impact time, Earth-like planets in the classical simulations acquire larger late-accreted masses than those in the Grand Tack simulations (see Fig. 1), because the planetesimal population is more dispersed in the classical scenario and therefore decays more slowly.

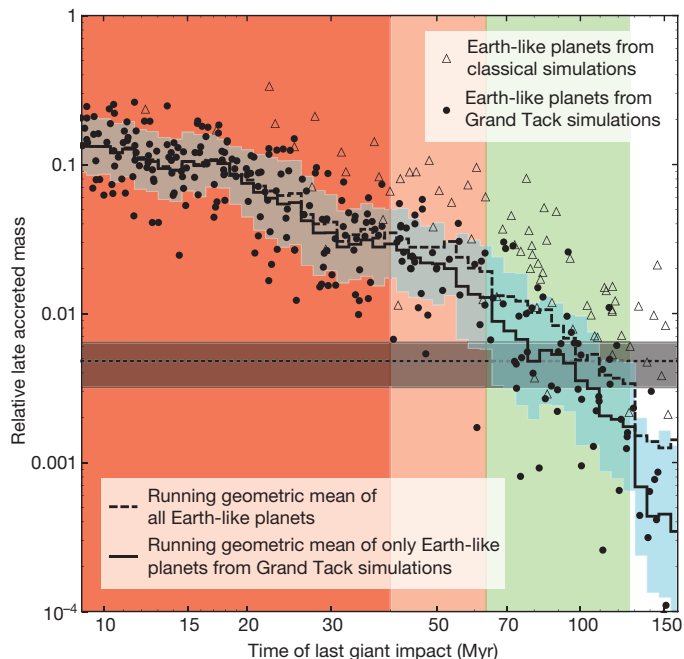


Figure 1 | The late-accreted mass relative to each synthetic Earth-like planet's final mass as a function of the time of the last giant impact.

Triangles represent Earth-like planets from the first category: classical simulations with Jupiter and Saturn near their contemporary orbits^{7,8}. Circles represent Earth-like planets from the second category: Grand Tack simulations with a truncated protoplanetary disk^{9,10}. The black line resembling a staircase is the moving geometric mean of the late-accreted masses in the Grand Tack simulations evaluated at logarithmic time intervals with a spacing parameter of 0.025 and a width parameter twice that. The blue region encloses the 1σ variance of the late-accreted mass, computed assuming that the latter is distributed log-normally about the geometric mean. Always predicting larger late-accreted masses for each last giant impact time, the dotted staircase is the geometric mean obtained by also considering the classical simulations, although those simulations do not fit Solar System constraints as well as the Grand Tack simulations do. The horizontal dashed line and enclosing darkened region are the best estimate and 1σ uncertainty of the late-accreted mass inferred from the HSE abundances in the mantle (chondritic mass): $4.8 \pm 1.6 \times 10^{-3} M_{\oplus}$. The best estimate for the intersection of the correlation and the chondritic mass is 95 ± 32 Myr after condensation. The dark and light red regions highlight Moon-formation times that are ruled out with 99.9% (40 Myr) and 85% (63 Myr) confidence or greater, respectively.

The correlation displayed in Fig. 1 can be used as a clock that is independent of radiometric dating systems. The late-accreted mass is input into this clock and the time of the last giant impact is read out. A traditional estimate for the late-accreted mass can be obtained from the highly siderophile element (HSE) abundances in Earth's mantle relative to the HSE abundances in chondritic meteorites^{11,12}. HSEs partition strongly into iron, and so are transported from the mantle to the core during core formation. In this process, the element ratios are strongly fractionated relative to chondritic proportions¹⁹. The HSEs in Earth's mantle are significantly depleted relative to chondritic bodies—a clear consequence of core formation—and yet the remaining HSEs are in chondritic or near-chondritic proportions relative to each other^{13,14}. This is commonly interpreted^{11,12,19,20} as evidence that all or a large portion of the HSEs currently in the mantle were delivered by chondritic bodies after the closure of Earth's core, an accretion phase known as the Late Veneer²⁰. To account for the observed mantle budget of HSEs, we estimate that a chondritic mass of $4.8 \pm 1.6 \times 10^{-3} M_{\oplus}$ is necessary, where M_{\oplus} represents an Earth mass. This mass does include contributions from the era known as the Late Heavy Bombardment. Current mass estimates for this very late (approximately 500 Myr after condensation) accretion are²¹ $10^{-3} M_{\oplus}$, which we added to the late-accreted masses of our synthetic Earth-like planets, but it only accounts for about 2% of the chondritic mass and therefore does not play an important part in our analysis of the correlation.

The chondritic mass can only be identical to the late-accreted mass or to the Late Veneer mass if the Moon-forming event stripped all of the HSEs from Earth's mantle or was the last episode of growth for Earth's core, respectively (as is traditionally assumed). However, these conditions are not necessarily true. Consider that some projectiles colliding with Earth after the Moon-forming event might have been differentiated, so that their HSEs were contained in their cores. If part of these cores had merged with Earth's core²², then the late-accreted mass would clearly be larger than the chondritic mass, because there would be no HSE record of this fraction of the projectile cores in Earth's mantle. Additionally, in this case, given that iron (and therefore HSEs) would have been simultaneously added to Earth's mantle and its core, the chondritic mass would be larger than the Late Veneer mass, which is geochemically defined as the mass accreted to Earth after the core has stopped growing.

In fact, as explained in detail in the Methods and in Extended Data Figs 3 and 4, it is unlikely that more than 50% of a projectile's core directly reaches Earth's core, otherwise geochemical models cannot reproduce the tungsten isotope composition of Earth's mantle²³. Moreover, a large late-accreted mass, delivered in only a few objects so as to explain the relative HSE abundances of Earth and Moon¹², would have left a detectable isotopic signature on Earth relative to the Moon^{24,25}. Thus, even when considering these more complex possibilities, geochemical evidence constrains the late-accreted mass probably not to exceed $0.01 M_{\oplus}$ (see Methods).

For these reasons, we first make the usual assumption that the late-accreted mass and the HSE-derived chondritic mass are identical. In this case, not a single simulated Earth-like planet with a last giant impact earlier than 48 Myr since condensation has a late-accreted mass in agreement with the value estimated from HSEs (see Fig. 1). Of those forming in less than 48 Myr, only one planet is near the upper 1σ bounds of the chondritic mass. Only after 67 Myr since condensation are there Earth-like planets with late-accreted masses consistently within the 1σ uncertainty bounds for the chondritic mass. After 126 Myr since condensation, the late-accreted masses of Earth-like planets are often significantly below the lower limit set by the HSE measurements.

We calculate the log-normal mean and standard deviation of the late-accreted masses of all Earth-like planets with last giant impacts within a range around a chosen time (see Fig. 1). We interpret these distributions as a model of the likelihood of a specific late-accreted mass given a last giant impact time. Given this likelihood model, we

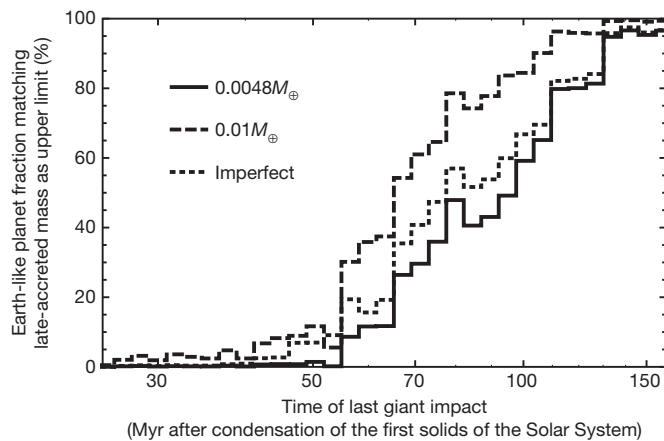


Figure 2 | The likelihood that a planet suffering a last giant impact within a specific range of times has a late-accreted mass less than or equal to the chondritic mass of $4.8 \pm 1.6 \times 10^{-3} M_{\oplus}$. The probability calculation uses the same bins as Fig. 1 but only includes the Grand Tack simulations, because they best reproduce the terrestrial planets^{9,10,18}. The solid line shows this probability assuming perfect accretion and corresponds exactly to the late-accreted masses shown as circles in Fig. 1. The lower 1σ limit for the Moon formation age is 63 Myr, which corresponds to a 15% probability that an Earth-like planet with a last giant impact at that age is characterized by a late-accreted mass equal or smaller than the chondritic mass. The dashed line shows the same calculation but for a late-accreted mass less than or equal to $0.01 M_{\oplus}$, which is an upper limit established from a number of elemental and isotopic constraints (see Supplementary Information). The dotted line shows the same calculation as for the solid line—that is, using a chondritic mass—but assuming imperfect accretion during collisions. This decreases the late-accreted masses by a variable amount depending on the impact characteristics of the late-accreted projectiles onto each planet²⁶. However, this calculation underestimates the late-accreted mass because a large fraction of the ejected material would be subsequently reaccreted^{27,28} and projectile core material is less likely to be ejected post-impact²⁹. Consequently, the dotted line overestimates the likelihood that a planet matches the chondritic mass constraint. A realistic estimate therefore lies between the solid and dotted curves (probably closer to the former).

compute the fraction of Earth-like planets with late-accreted masses not exceeding the HSE-derived chondritic mass, also taking into account the uncertainties on the latter. Using only the Grand Tack simulations, which provide the best match to the terrestrial planets^{9,10,18}, the probability that an Earth-like planet with a last giant impact at or before 40 Myr since condensation has an HSE budget consistent with observations is 0.1% or less (see Fig. 2). Furthermore, if we relax the assumption that the late-accreted mass is equal to the chondritic mass, but consider that the former can be up to $0.01 M_{\oplus}$, then a Moon-forming event at or before 40 Myr since condensation is still ruled out at a 97.5% confidence level or higher.

Before accepting the proposed HSE clock, a couple of caveats need to be discussed. First, our simulations always assume perfect accretion—that is, all mass from both colliding bodies ends up in the remaining planet. Accounting for imperfect accretion—in which a fraction of the total mass is ejected away from the planet, the true late-accreted mass is probably smaller than that inferred from our simulations. Using the characteristics of each impact (for example, impact velocity and angle) recorded in our simulations and an algorithm developed from a large suite of numerical experiments²⁶, we estimate the fraction of the projectile mass retained in each impact. The effect of imperfect accretion somewhat increases the likelihood of early giant impacts (see Fig. 2). However, *N*-body simulations that have incorporated imperfect accretion showed that most of the mass not retained in the impacts is subsequently re-accreted^{27,28}. Thus, our calculation incorporating imperfect accretion should be interpreted as an upper limit to the likelihood for a given last-giant-impact time. Reinforcing this analysis, if the HSEs are

mostly delivered from the cores of differentiated projectiles as in the scenario of Bottke *et al.*¹², imperfect accretion has a minimal effect on the HSE budget because the material that is lost into space comes predominantly from the projectile's mantle²⁹.

Second, mutually catastrophic collisions between planetesimals may break them into ever smaller pieces until they are small enough to be removed by solar radiation before they can be accreted onto planets. If this were a significant process, the late-accreted mass would be smaller than estimated from our simulations, which do not include this process. However, the size distribution of craters on the lunar highlands suggests that—as in the current asteroid belt—most of the planetesimals' mass was in objects larger than 100 km in diameter³⁰. The collisional comminution for bodies this large is negligible. Moreover, most of the late-accreted mass was probably delivered by Ceres-sized (about 1,000 km across) or even larger bodies to explain the relative difference in HSE abundances between Earth and the Moon¹².

Considering all of the above, we argue that the clock derived from the correlation between late-accreted mass and last-giant-impact age is robust and may be the most reliable way to estimate the age of the Moon-forming event. Given the current constraints on Earth's late-accreted mass, it establishes that this event occurred significantly later than 40 Myr and probably 95 ± 32 Myr after the formation of the first Solar System solids, in agreement with some^{5,6}, but not all, previous estimates based on radiometric chronometers. Future analyses will establish firmer constraints on the late-accreted mass, for instance, by determining the difference in W-isotope composition between Earth and the Moon more precisely. From this, we can obtain better limits on the timing of the Moon-forming event through our clock.

Our analysis of numerical simulations can also be used to support or invalidate different accretion scenarios. For instance, Albaredo *et al.*²² consider the possibility of the impact with Earth of a few projectiles that have a total mass of $0.04 M_{\oplus}$ as late as 130 Myr since condensation, but Fig. 1 shows that it is extremely unlikely for so much mass to be delivered so late.

A late Moon formation has at least two profound implications. First, it constrains the dynamical conditions of the disk from which the planets accreted and the physical properties of the disk material. For instance, a late last giant impact implies that most of the mass was in the embryo population rather than the planetesimal population. Second, reconciling a late Moon-forming event with radiometric chronometers that suggest the opposite result may require challenging fundamental assumptions, such as envisioning a Moon-forming event that did not reset all clocks simultaneously and left significant parts of the mantle non-equilibrated with the core. This may argue in favour of some of the new scenarios proposed for the Moon-forming collision^{31,32}, which distribute impact energy heterogeneously and may leave a significant portion of Earth's mantle relatively undisturbed.

METHODS SUMMARY

The *N*-body simulations come from published sources^{7–10} and original calculations. All simulations begin with a protoplanetary disk of around 1,000–2,000 Ceres-mass planetesimals and around 100 Moon-mass to Mars-mass embryos extending from an inner edge at 0.3–0.7 AU from the Sun to an outer edge at 4.0–4.5 AU. Classical simulations assume that Jupiter and Saturn form on or near their current orbits^{7,8} and begin when the gas in the disk disappears. Grand Tack simulations begin slightly earlier than the classical simulations, about 0.6 Myr before the gas disperses. During this short gas phase, Jupiter and Saturn migrate inward to 1.5 AU and 2 AU, respectively, and then outward through the disk according to the simplest Grand Tack scheme tested by Walsh *et al.*⁹ (“Saturn's core growing in the 2:3 resonance with Jupiter.”). After 150 Myr since the start of the simulations, Earth-like planets in each system were extracted and used in the analysis presented above.

The canonical estimate of the late-accreted mass from the HSE abundances of Earth's mantle come from calculations using measurements of HSEs in Earth's mantle¹³ and a careful compilation of HSE chondrite measurements¹⁴. This estimate can be exceeded if projectiles do not deliver HSEs in chondritic abundances to Earth's mantle. Despite this, we show that the late-accreted mass is probably less than $0.01 M_{\oplus}$ for three different sets of isotopic constraints: (1) the Hf–W system

limits the amount of projectile core material entering Earth's core without emulsification and interaction with Earth's mantle²³; (2) the O and Ti isotopic similarities of Earth and Moon restrict the contributions from different meteoritic compositions; and (3) the measured W-isotopic difference between Earth and Moon is controlled by the mass of differentiated planetesimals delivered during late accretion. Although each individual constraint is not airtight, the combination is such that the late-accreted mass is not likely to exceed $0.01M_{\oplus}$ unless very contrived scenarios are considered.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 16 September 2013; accepted 17 February 2014.

- Yin, Q. *et al.* A short timescale for terrestrial planet formation from Hf–W chronometry of meteorites. *Nature* **418**, 949–952 (2002).
- Jacobsen, S. B. The Hf–W isotopic system and the origin of the Earth and Moon. *Annu. Rev. Earth Planet. Sci.* **33**, 531–570 (2005).
- Taylor, D. J., McKeegan, K. D. & Harrison, T. M. Lu–Hf zircon evidence for rapid lunar differentiation. *Earth Planet. Sci. Lett.* **279**, 157–164 (2009).
- Touboul, M., Kleine, T., Bourdon, B., Palme, H. & Wieler, R. Late formation and prolonged differentiation of the Moon inferred from W isotopes in lunar metals. *Nature* **450**, 1206–1209 (2007).
- Allège, C. J., Manhès, G. & Göpel, C. The major differentiation of the Earth at 4.45 Ga. *Earth Planet. Sci. Lett.* **267**, 386–398 (2008).
- Halliday, A. N. A young Moon-forming giant impact at 70–110 million years accompanied by late-stage mixing, core formation and degassing of the Earth. *Phil. Trans. R. Soc. A* **366**, 4163–4181 (2008).
- O'Brien, D. P., Morbidelli, A. & Levison, H. F. Terrestrial planet formation with strong dynamical friction. *Icarus* **184**, 39–58 (2006).
- Raymond, S. N., O'Brien, D. P., Morbidelli, A. & Kaib, N. A. Building the terrestrial planets: constrained accretion in the inner Solar System. *Icarus* **203**, 644–662 (2009).
- Walsh, K. J., Morbidelli, A., Raymond, S. N., O'Brien, D. P. & Mandell, A. M. A low mass for Mars from Jupiter's early gas-driven migration. *Nature* **475**, 206–209 (2011).
- O'Brien, D. P., Walsh, K. J., Morbidelli, A., Raymond, S. N. & Mandell, A. M. Water delivery and giant impacts in the 'Grand Tack' scenario. *Icarus* (submitted).
- Chyba, C. F. Terrestrial mantle siderophiles and the lunar impact record. *Icarus* **92**, 217–233 (1991).
- Bottke, W. F., Walker, R. J., Day, J. M. D., Nesvorný, D. & Elkins-Tanton, L. Stochastic Late Accretion to Earth, the Moon, and Mars. *Science* **330**, 1527–1530 (2010).
- Becker, H. *et al.* Highly siderophile element composition of the Earth's primitive upper mantle: constraints from new data on peridotite massifs and xenoliths. *Geochim. Cosmochim. Acta* **70**, 4528–4550 (2006).
- Walker, R. J. Highly siderophile elements in the Earth, Moon and Mars: update and implications for planetary accretion and differentiation. *Chemie Erde Geochem.* **69**, 101–125 (2009).
- König, S. *et al.* The Earth's tungsten budget during mantle melting and crust formation. *Geochim. Cosmochim. Acta* **75**, 2119–2136 (2011).
- Wetherill, G. W. Why isn't Mars as big as Earth? *Abstr. Lunar Planet. Sci. Conf.* **22**, 1495 (1991).
- Chambers, J. E. & Cassen, P. M. The effects of nebula surface density profile and giant-planet eccentricities on planetary accretion in the inner solar system. *Meteorit. Planet. Sci.* **37**, 1523–1540 (2002).
- Hansen, B. M. S. Formation of the terrestrial planets from a narrow annulus. *Astrophys. J.* **703**, 1131–1140 (2009).
- Mann, U., Frost, D. J., Rubie, D. C., Becker, H. & Audétat, A. Partitioning of Ru, Rh, Pd, Re, Ir and Pt between liquid metal and silicate at high pressures and high temperatures—implications for the origin of highly siderophile element concentrations in the Earth's mantle. *Geochim. Cosmochim. Acta* **84**, 593–613 (2012).
- Chou, C. L. Fractionation of siderophile elements in the Earth's upper mantle. *Abstr. Lunar Planet. Sci. Conf.* **9**, 219–230 (1978).
- Morbidelli, A., Marchi, S., Bottke, W. F. & Kring, D. A. A sawtooth-like timeline for the first billion years of lunar bombardment. *Earth Planet. Sci. Lett.* **355**, 144–151 (2012).
- Albarede, F. *et al.* Asteroidal impacts and the origin of terrestrial and lunar volatiles. *Icarus* **222**, 44–52 (2013).
- Rudge, J. F., Kleine, T. & Bourdon, B. Broad bounds on Earth's accretion and core formation constrained by geochemical models. *Nature Geosci.* **3**, 439–443 (2010).
- Wiechert, U. *et al.* Oxygen isotopes and the Moon-forming giant impact. *Science* **294**, 345–348 (2001).
- Zhang, J., Dauphas, N., Davis, A. M., Leya, I. & Fedkin, A. The proto-Earth as a significant source of lunar material. *Nature Geosci.* **5**, 251–255 (2012).
- Leinhardt, Z. M. & Stewart, S. T. Collisions between gravity-dominated bodies. I. Outcome regimes and scaling laws. *Astrophys. J.* **745**, 79 (2012).
- Kokubo, E. & Genda, H. Formation of terrestrial planets from protoplanets under a realistic accretion condition. *Astrophys. J.* **714**, L21–L25 (2010).
- Chambers, J. E. Late-stage planetary accretion including hit-and-run collisions and fragmentation. *Icarus* **224**, 43–56 (2013).
- Asphaug, E., Agnor, C. B. & Williams, Q. Hit-and-run planetary collisions. *Nature* **439**, 155–160 (2006).
- Strom, R. G., Malhotra, R., Ito, T., Yoshida, F. & Kring, D. A. The origin of planetary impactors in the inner Solar System. *Science* **309**, 1847–1850 (2005).
- Čuk, M. & Stewart, S. T. Making the Moon from a fast-spinning Earth: a giant impact followed by resonant despinning. *Science* **338**, 1047–1052 (2012).
- Reufer, A., Meier, M. M. M., Benz, W. & Wieler, R. A hit-and-run giant impact scenario. *Icarus* **221**, 296–299 (2012).

Acknowledgements S.A.J., A.M., D.P.O'B. and D.C.R. were supported by the European Research Council Advanced Grant 'ACCRETE' (contract number 290568). D.P.O'B. was also supported by grant NNX09AE36G from NASA's Planetary Geology and Geophysics research programme. S.N.R. acknowledges support from the NASA Astrobiology Institute's Virtual Planetary Laboratory lead team, supported by NASA under cooperative agreement NNNH05ZDA001C.

Author Contributions A.M., K.J.W., S.N.R. and D.P.O'B. initiated a study of numerically simulating terrestrial planet formation. A.M., S.A.J., S.N.R., D.P.O'B. and K.J.W. ran *N*-body simulations and provided reduced data. S.A.J. analysed results. S.A.J., A.M. and D.C.R. discussed the meaning of the discovered relationship. S.A.J. wrote the paper with guidance from A.M. All authors commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.A.J. (seth.jacobson@oca.eu).

The hippocampal CA2 region is essential for social memory

Frederick L. Hitti¹ & Steven A. Siegelbaum^{1,2}

The hippocampus is critical for encoding declarative memory, our repository of knowledge of who, what, where and when¹. Mnemonic information is processed in the hippocampus through several parallel routes involving distinct subregions. In the classic trisynaptic pathway, information proceeds from entorhinal cortex (EC) to dentate gyrus to CA3 and then to CA1, the main hippocampal output². Genetic lesions of EC (ref. 3) and hippocampal dentate gyrus (ref. 4), CA3 (ref. 5) and CA1 (ref. 6) regions have revealed their distinct functions in learning and memory. In contrast, little is known about the role of CA2, a relatively small area interposed between CA3 and CA1 that forms the nexus of a powerful disynaptic circuit linking EC input with CA1 output⁷. Here we report a novel transgenic mouse line that enabled us to selectively examine the synaptic connections and behavioural role of the CA2 region in adult mice. Genetically targeted inactivation of CA2 pyramidal neurons caused a pronounced loss of social memory—the ability of an animal to remember a conspecific—with no change in sociability or several other hippocampus-dependent behaviours, including spatial and contextual memory. These behavioural and anatomical results thus reveal CA2 as a critical hub of sociocognitive memory processing.

Although the CA2 region was first described in 1934 (ref. 8), relatively little is known about its functional properties and behavioural role. To examine the importance of this region, we generated a transgenic mouse line (*Amigo2-Cre*) that expresses Cre recombinase predominantly in CA2 pyramidal neurons (PNs) in adult mice (Extended Data Fig. 1). Because this line expresses Cre throughout the brain during early development, as well as in certain limited areas outside CA2 in the adult, we stereotactically injected Cre-dependent adeno-associated virus (AAV) into the hippocampus of adult *Amigo2-Cre* mice to limit viral expression to CA2 pyramidal cells.

To determine the specificity of CA2 expression in the transgenic line, we bilaterally injected into dorsal hippocampus a Cre-dependent AAV to express yellow fluorescent protein (YFP) in Cre⁺ cells (Fig. 1a). We observed selective and robust YFP expression in CA2 PNs throughout dorsal hippocampus^{9–11} (Fig. 1b and Extended Data Fig. 2a). We confirmed that the Cre⁺ cells were indeed CA2 PNs by demonstrating co-staining for RGS14 (ref. 12) ($97.38 \pm 0.31\%$ overlap; $n = 4$ mice, 2,546 cells; Fig. 1c–e and Extended Data Fig. 3) and other known CA2 PN markers (Extended Data Fig. 2). In contrast, there was no co-staining for a CA1 PN marker (Extended Data Fig. 2). In addition, the electrophysiological properties of the YFP⁺ neurons differed significantly from those of CA1 PNs (Extended Data Table 1) and largely matched the values previously reported for CA2 pyramidal neurons⁷. Only a minute fraction of YFP⁺ neurons were also GABA⁺ ($0.16 \pm 0.16\%$; $n = 3$ mice, 1,539 cells), showing the specific targeting of CA2 excitatory PNs (Fig. 1f, g and Extended Data Fig. 3). Finally, our AAV injections resulted in the targeting of the vast majority of CA2 PNs in the dorsal hippocampus, measured by the percentage of RGS14⁺ cells that were also YFP⁺ ($82.33 \pm 2.37\%$, $n = 4$ mice, 2,992 cells).

Next we mapped CA2 synaptic input and output by using viral tracing strategies that take advantage of the genetic targeting of CA2 PNs in the

Amigo2-Cre mice, and largely confirmed results of previous studies using conventional¹³ and genetic-based¹⁴ approaches. Monosynaptic inputs to CA2 PNs were determined by trans-synaptic retrograde labelling using an EnvA pseudotyped ΔG rabies virus¹⁵ (Extended Data Fig. 4). Unilateral viral injections revealed bilateral inputs from CA3 and CA2 (Fig. 2a, b) and strong unilateral input from both lateral and medial EC layer II neurons (Fig. 2c, d). In addition, synaptic inputs were detected from medial septum and diagonal band (Fig. 2e), median raphe nucleus (Fig. 2f), and the supramammillary nucleus of the hypothalamus (Fig. 2g).

We observed only sparse labelling of EC layer III neurons with the rabies virus approach. Our laboratory had previously concluded that EC LIII axons provide strong excitatory drive to CA2 PNs, on the basis of the finding that large excitatory postsynaptic potentials are evoked in CA2 PNs with a focal stimulating electrode placed in the stratum lacunosum of the CA1 region⁷, where axons from LIII EC neurons are thought to provide the predominant source of excitatory inputs. Our present results, combined with recent results^{13,14}, suggest that these synaptic responses recorded in CA2 PNs may result from activation of LII fibres that course through or near the stratum lacunosum in CA1.

Output projections from CA2 were determined by expressing YFP in CA2 PNs (as in Fig. 1) and examining brains for YFP-fluorescent axons. Unilateral viral injections resulted in strong bilateral labelling in hippocampal CA1, CA2 and CA3 regions, with densest projections observed in stratum oriens and weaker projections detected in stratum radiatum (Fig. 2h, i). We did not observe extra-hippocampal outputs.

These anatomical results generally support previous^{13,14} findings. However, we failed to observe vasopressinergic input to CA2 from the paraventricular nucleus of the hypothalamus¹³, which may reflect an inability of the trans-synaptic rabies tracing system to label peptidergic inputs¹⁵. In addition, we did not observe CA2 output to the supramammillary nucleus as reported previously using conventional tracing methods¹³. We surmise that this output may represent an inhibitory projection from CA2 because our technique selectively labelled PNs. Finally, we did not observe CA2 output to EC layer II (ref. 16), perhaps because the anterograde tracing failed to detect weak connections.

To examine directly the functional and behavioural relevance of CA2, we used the *Amigo2-Cre* mouse line to inactivate output from CA2 PNs selectively. We injected into the dorsal hippocampus of the *Amigo2-Cre* mice a Cre-dependent AAV to express tetanus neurotoxin (TeNT) light chain fused to enhanced green fluorescent protein (eGFP–TeNT) in CA2 PNs to block their synaptic output. We first verified the efficacy of this approach and characterized the influence of CA2 on its CA1 PN targets by using Cre-dependent AAVs to co-express the light-activated cation channel channelrhodopsin-2 (ChR2)¹⁷ with either TeNT or YFP. Low-intensity illumination (using 2-ms pulses of 470-nm light at 3 mW mm^{-2}) focused on CA2 reliably triggered action potentials in CA2 PNs, as seen by the presence of fast action currents in cell-attached patch clamp recordings (Fig. 3a–c). Similar rates of spiking were seen in neurons that co-expressed either YFP (Fig. 3b) or TeNT (Fig. 3c) with ChR2, indicating that the TeNT did not inhibit excitability.

¹Department of Neuroscience, Kavli Institute, College of Physicians and Surgeons, Columbia University 1051 Riverside Drive, New York, New York 10032, USA. ²Department of Pharmacology, Howard Hughes Medical Institute, College of Physicians and Surgeons, Columbia University 1051 Riverside Drive, New York, New York 10032, USA.

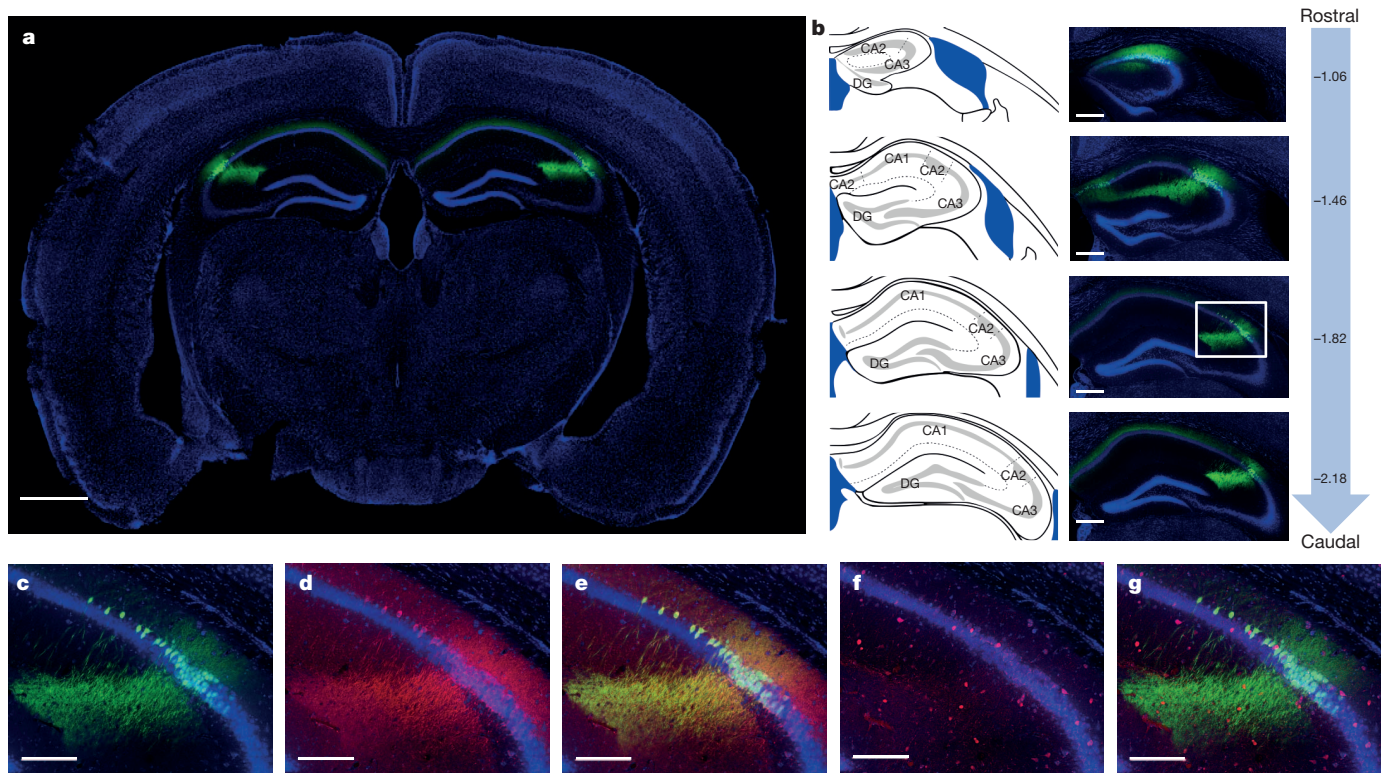


Figure 1 | Genetic targeting of the CA2 subfield using the *Amigo2-Cre* mouse line. **a**, Bilateral hippocampal injection of Cre-dependent YFP AAV in *Amigo2-Cre* mice resulted in specific expression of YFP (green) in CA2 PNs ($n = 64$ mice). **b**, Extent of transduction. Left: adapted reference atlas images⁹. Centre: YFP expression. Right: distance (in mm) from bregma along the rostrocaudal axis. DG, dentate gyrus. **c–g**, Magnified images of the boxed area

in **b**, **c**, YFP (green). **d**, RGS14 staining (red, $n = 4$ mice). **e**, Merge of **c** and **d**, showing YFP and RGS14 overlap. **f**, GABA staining (red, $n = 3$ mice). **g**, Merge of **c** and **f**, showing no GABA and YFP overlap. Panels show coronal sections with Nissl counterstain (blue). Scale bars, 1,000 μm (**a**), 400 μm (**b**), 200 μm (**c–g**).

Next we determined the strength of synaptic transmission from CA2 to CA1 PNs by using whole-cell current-clamp recordings to measure light-evoked postsynaptic potentials (PSPs) in CA1 PNs from hippocampal slices in which Chr2 and YFP were expressed in CA2 PNs (Fig. 3d). In agreement with anatomical mapping (Fig. 2h, i) and paired recordings⁷, focal photostimulation delivered to CA1 stratum

oriens and stratum radiatum regions evoked robust monosynaptic PSPs (mean latency 1.22 ± 0.06 ms, $n = 119$ observations) in nearby CA1 PNs (Fig. 3e). Increasing the light intensity recruited progressively larger PSPs, presumably as a result of an increase in the number of optically activated CA2 axons (Fig. 3e, f). In stark contrast, in slices in which TeNT was co-expressed with Chr2 in CA2 PNs, illumination over a wide range of intensities produced little or no synaptic response in CA1 neurons (Fig. 3e, f), showing the efficacy of the TeNT lesion.

What are the behavioural consequences of inactivation of CA2? To address this question we compared the behaviour of control mice (CA2-YFP) with mice in which CA2 PNs were inactivated (CA2-TeNT), using viral injections in dorsal hippocampus¹¹. Functional inactivation of dorsal CA2 did not alter locomotor activity or anxiety-like behaviour (Extended Data Fig. 5). Inactivation of CA2 also did not significantly alter hippocampus-dependent spatial memory assessed by the Morris water maze (although there was a trend for the CA2-inactivated mice to learn the task more slowly; Extended Data Fig. 6). Nor was there any

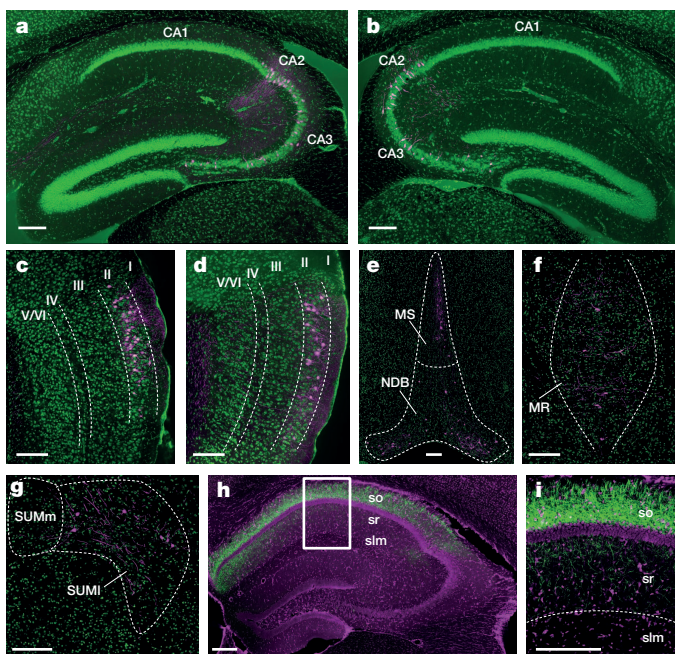


Figure 2 | Genetically targeted tracing of the CA2 circuit. **a–g**, Monosynaptic inputs to CA2 revealed with pseudotyped rabies virus ($n = 8$ mice). Cells labelled with rabies appear magenta; Nissl stain shown in green. **a–d**, Sagittal sections; **e–g**, coronal sections. **a**, **b**, Labelled neurons in CA2 and CA3 ipsilateral (**a**) and contralateral (**b**) to the hemisphere of rabies virus injection. Rabies labelling shows monosynaptic inputs from lateral EC (**c**), medial EC (**d**), medial septum (MS) and nucleus of the diagonal band (NDB) (**e**), median raphe (MR) (**f**) and lateral supramammillary nucleus (SUMl) (**g**). Fluorescent processes in (**c**, **d**) may represent dendritic or axonal labelling. SUMm, medial supramammillary nucleus. **h**, Output of CA2 revealed by axonal YFP signal (green, $n = 6$ mice). Nissl stain (magenta). so, stratum oriens; sr, stratum radiatum; slm, stratum lacunosum-moleculare. **i**, Magnification of boxed area in **h**. Note that CA2 makes strong projections to stratum oriens and weaker projections to stratum radiatum of CA1. Scale bars, 200 μm .

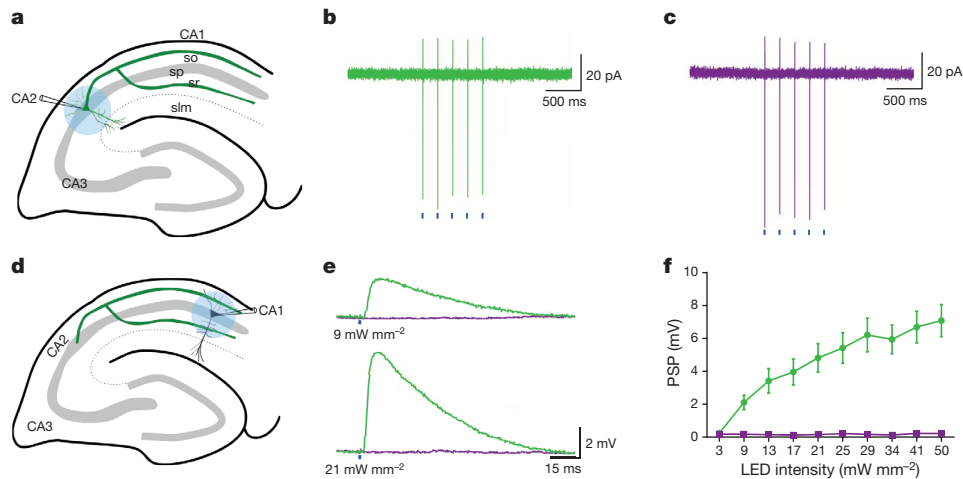


Figure 3 | Electrophysiological verification of CA2 inactivation with TeNT. **a**, Experimental setup for photostimulation of CA2 PNs. sp, stratum pyramidale. **b**, **c**, Action currents recorded from CA2 PNs expressing YFP and Chr2 ($n = 6$ neurons) (**b**) or TeNT and Chr2 ($n = 4$ neurons) (**c**) in response to five 2-ms blue (470-nm) light pulses (blue bars). **d**, Experimental setup for whole-cell current-clamp recordings of photostimulated PSPs in CA1 PNs.

change in hippocampus-dependent contextual fear memory or amygdala-dependent auditory fear memory (Extended Data Fig. 7).

The finding that CA2 PNs integrate synaptic input from lateral EC (which conveys non-spatial information¹⁸) with subcortical input from both the serotonergic median raphe nucleus¹⁹ and the hypothalamic supramammillary nucleus²⁰ suggests a potential role for CA2 in non-spatial hippocampal tasks. Previous studies have shown that the messenger RNA for the vasopressin 1b receptor (*Avpr1b*) is strongly expressed in CA2 (ref. 21) and that unconditional deletion of this gene impairs social recognition memory^{22,23}. However, *Avpr1b* mRNA is also expressed outside hippocampus²¹, and its deletion results in changes in non-hippocampus-dependent behaviours, including reduced aggression and decreased sociability^{22,23}, raising questions as to the selective role of CA2 in the knockout phenotype²⁴.

To assess directly the role of CA2 in social behaviour, we first compared the performance of CA2-YFP with that of CA2-TeNT mice in a three-chamber test of sociability²³, which examines the normal preference of a subject mouse for a chamber containing a littermate versus an empty chamber (Fig. 4a). In contrast to the effect of *Avpr1b* deletion, selective silencing of CA2 did not alter sociability as the CA2-YFP and CA2-TeNT groups displayed a significant and similar preference for the compartment containing the littermate (Fig. 4a).

In contrast to their normal sociability, CA2-TeNT mice displayed a profound deficit in social recognition as determined by a three-chamber social novelty test²³ (Fig. 4b). In this test, social recognition was measured by the increased time that a subject mouse spent interacting with a novel unrelated mouse compared with the time it spent interacting with a familiar co-housed littermate. Multiple comparison testing revealed that the CA2-YFP control group demonstrated a significant preference for the compartment containing the novel animal, whereas the CA2-TeNT group did not (Fig. 4b). Moreover, the difference score (time spent exploring the novel mouse minus the time spent exploring the familiar mouse) of the CA2-TeNT group was significantly less than that of the CA2-YFP group (Fig. 4b). This deficit was not due to a lack of interest in novelty as such, because the CA2-TeNT mice showed a normal preference for a novel object as assayed by two different novel-object-recognition protocols (Extended Data Fig. 8).

Because the social novelty test does not incorporate a defined learning phase or delay period, we next conducted a more specific test of social memory, the direct interaction test²⁵. For this test a subject mouse was exposed to an unfamiliar mouse in trial 1. After a 1-h inter-trial interval, the subject mouse was either re-exposed to the same mouse as

e, PSPs recorded when YFP ($n = 14$ neurons, green) or TeNT ($n = 14$ neurons, magenta) was co-expressed with Chr2 in CA2 PNs. **f**, Mean input-output curve of PSP as a function of light intensity when YFP (green) or TeNT (magenta) was co-expressed with Chr2 in CA2 PNs. Results are means \pm s.e.m.

that encountered in trial 1 (Fig. 4c) or exposed to a second, unfamiliar, mouse (Fig. 4d). Social memory, measured as the decreased time that a subject mouse spends exploring a previously encountered mouse, was fully suppressed by CA2 inactivation (Fig. 4c). In contrast, CA2 silencing did not alter sociability, because CA2-YFP and CA2-TeNT subject mice showed similar and unchanging exploration times during trials 1 and 2 when two different unfamiliar mice were encountered in the two trials (Fig. 4d).

We next conducted a more stringent five-trial social memory assay²⁶ to confirm that CA2-inactivation abolishes social memory. In this assay, a stimulus mouse was presented to a subject mouse for four successive trials. On the fifth trial, a novel stimulus mouse was introduced (Fig. 4e). The CA2-YFP control group displayed normal social memory, as demonstrated by a marked habituation (decreased exploration) during the first four trials and a striking dishabituation (increased exploration) on presentation of a novel animal on the fifth trial. In contrast, the CA2-TeNT group showed no significant habituation during the four exposures to the stimulus mouse or dishabituation to the novel stimulus mouse, thus confirming the necessity of CA2 for social memory.

Because olfaction is crucial for normal social interaction²⁷, we examined whether CA2 silencing influenced the detection or recognition of non-social or social odours. CA2-TeNT mice showed no loss in the ability to detect the presence of food buried under a deep layer of cage bedding, a test of non-social odour detection (Extended Data Fig. 9a). Next we used an olfactory habituation/dishabituation test (Extended Data Fig. 9b) and found that CA2 inactivation also had no effect on the ability of mice to detect or discriminate either non-social or social odours. We therefore conclude that the deficit in social memory in the CA2-TeNT mice was not due to a defect in sensing social or non-social odours.

In this study we developed and validated an *Amigo2-Cre* mouse line that enables the precise genetic targeting of excitatory CA2 PNs, allowing us to map selectively the inputs and outputs of this largely unexplored region and demonstrate that the CA2 subfield is essential for social memory. Although we observed a fairly specific deficit in social memory after inactivation of dorsal CA2 pyramidal neurons, our results do not rule out the possibility that CA2 may participate more generally in hippocampus-dependent memory tasks. Thus, other regions of hippocampus may be able to compensate for the loss of any role that CA2 may normally have in performance of the water maze or contextual fear-conditioning tasks. Alternatively, CA2 may be selectively required for the performance of more demanding non-social memory tasks.

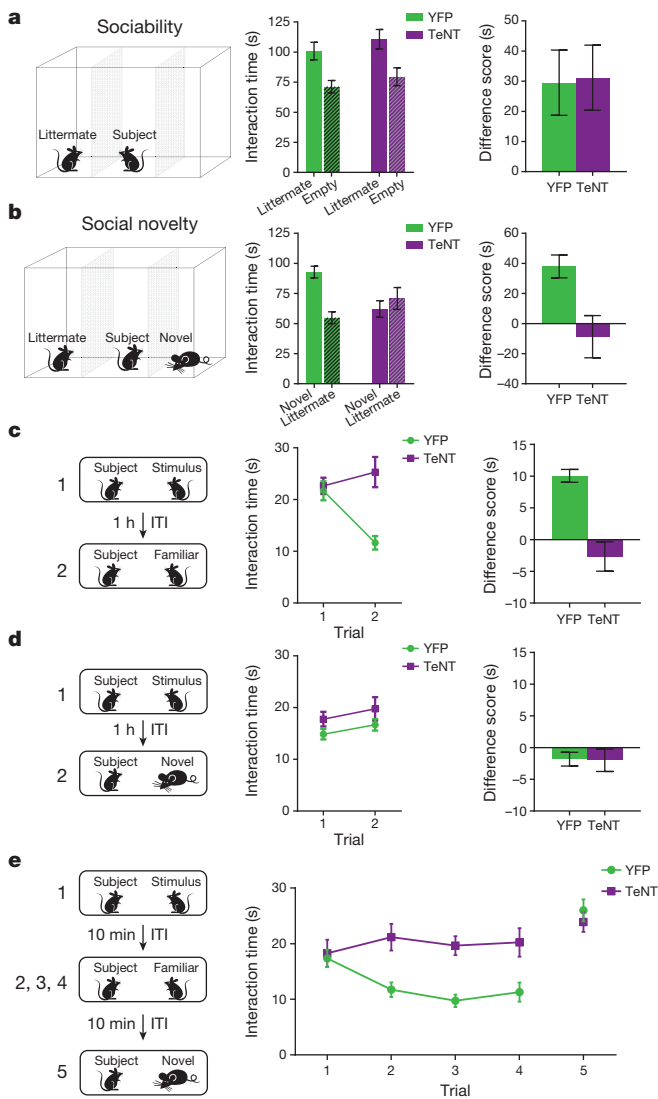


Figure 4 | Inactivation of CA2 impairs social memory. **a**, Left: sociability test. Middle: YFP ($n = 11$) and TeNT ($n = 13$) mice preferred the littermate chamber (YFP, $P = 0.0083$; TeNT, $P = 0.0055$; multiplicity-adjusted P values), with no significant difference in interaction times (two-way analysis of variance (ANOVA): treatment \times chamber $F(1,44) = 0.013$, $P = 0.91$; treatment $F(1,44) = 1.566$, $P = 0.22$; chamber $F(1,44) = 17.49$, $P = 0.0001$). Right: interaction time difference scores (littermate minus empty) were similar ($P = 0.9154$, two-tailed Student's t -test). **b**, Left: social novelty test. Middle: only YFP mice preferred the novel animal over a littermate (YFP, $P = 0.0012$; TeNT, $P = 0.3593$; multiplicity-adjusted P values); the groups differed significantly (ANOVA: treatment \times chamber $F(1,44) = 11.25$, $P = 0.0016$). Right: difference score (novel minus littermate) of TeNT group was significantly less than that of YFP group ($P = 0.0109$, two-tailed Student's t -test). **c**, **d**, Left: direct interaction test using identical (c) or different (d) stimulus animals in the two trials. **c**, Middle: only YFP mice displayed decreased investigation during trial 2 of mouse encountered in trial 1 (YFP, $n = 15$, $P < 0.0001$; TeNT, $n = 16$, $P = 0.1499$; multiplicity-adjusted P values); the groups differed significantly (ANOVA: treatment \times trial $F(1,29) = 24.23$, $P < 0.0001$). Right: TeNT group difference score (trial 1 minus trial 2) was less than that of YFP group ($P < 0.0001$; two-tailed Student's t -test). **d**, Middle: the two groups explored two different stimulus animals similarly (ANOVA: treatment \times trial $F(1,29) = 0.0068$, $P = 0.93$; treatment $F(1,29) = 2.405$, $P = 0.13$; trial $F(1,29) = 3.278$, $P = 0.0806$), with similar difference scores (right, $P = 0.93$, two-tailed Student's t -test). **e**, Five-trial social memory assay. YFP ($n = 15$), but not TeNT ($n = 14$), mice habituated to same mouse (trials 1–4) and dishabituated to a novel mouse (trial 5). The groups differed significantly (ANOVA: treatment \times trial $F(4,108) = 7.26$, $P < 0.0001$; treatment $F(1,27) = 7.86$, $P = 0.009$; trial $F(4,108) = 15.41$, $P < 0.0001$). ITI, inter-trial interval. Results are means \pm s.e.m.

The importance of human hippocampus for social memory is famously illustrated by the case of Henry Molaison (patient H.M.), who, after bilateral medial temporal lobe ablation, could not form new memories of people he had worked with for years²⁸. Lesions limited to the hippocampus also impair social memory in both humans¹ and rodents²⁵. Because several neuropsychiatric disorders are associated with altered social endophenotypes, our findings raise the possibility that CA2 dysfunction may contribute to these behavioural changes. This possibility is supported by findings of a decreased number of CA2 inhibitory neurons in individuals with schizophrenia and bipolar disorder²⁹, and altered vasopressin signalling in autism³⁰. Thus, CA2 may provide a new target for therapeutic approaches to the treatment of social disorders.

METHODS SUMMARY

Generation of *Amigo2-Cre* mouse line. A bacterial artificial chromosome (BAC) was modified to insert a Cre-HSV-polyA cassette at the translational start of *Amigo2*. Six B6CBA/F2 founders were generated. One line that selectively expressed Cre in CA2 of adult mice was backcrossed to C57BL/6J a minimum of six times and used in these experiments.

Subjects. Cre⁺ males were bred to C57BL/6J females to keep the *Amigo2-Cre* line hemizygous on the C57BL/6J background. Only Cre⁺ males were used for experiments, which were conducted 2–4 weeks after viral injection and approved by the Columbia University Institutional Animal Care and Use Committee.

Viruses. AAV5-EF1 α -FLEX-eYFP-WPRE-hGH, AAV5-EF1 α -FLEX-mCherry-WPRE-hGH and AAV5-CAG-FLEX-rabiesG-WPRE-hGH were obtained from the University of North Carolina vector core. (EnvA)SAD- Δ G-mCherry was produced at Columbia. AAV5-EF1 α -FLEX-hChR2(H134R)-EYFP-WPRE and AAV5-EF1 α -FLEX-eGFP-TeNT-WPRE-hGH were obtained from the UPenn vector core.

Stereotaxic injection. Male mice (more than 8 weeks old) were anaesthetized with isoflurane, and virus was injected into the dorsal hippocampus at -1.6 mm anteroposterior, ± 1.6 mm mediolateral, and -1.7 mm dorsoventral relative to bregma.

Electrophysiology. Two to three weeks after AAV injection, hippocampal slices 400 μ m thick were prepared. Blue light was delivered through a 20 \times objective to activate ChR2. Patch membrane voltage was held at -70 mV for cell-attached voltage-clamp recordings. Whole-cell recordings were obtained from CA1 PN in current-clamp mode with membrane at initial resting potential.

Behavioural tests. Two to five male *Amigo2-Cre* mice were housed per cage with *ad libitum* access to food and water. Tests were conducted during the light cycle. Half of the mice in each cage were injected with the YFP virus, and the other half were injected with the TeNT virus. The experimenter was blind to the group identities. Open field activity was recorded with Activity Monitor. The direct interaction, five-trial social memory and olfactory tests were scored online by the experimenter. All other tests were scored automatically by ANY-maze.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 5 December 2013; accepted 10 January 2014.

Published online 23 February 2014.

- Squire, L. R. & Wixted, J. T. The cognitive neuroscience of human memory since H.M. *Annu. Rev. Neurosci.* **34**, 259–288 (2011).
- van Strien, N. M., Cappaert, N. L. & Witter, M. P. The anatomy of memory: an interactive overview of the parahippocampal–hippocampal network. *Nature Rev. Neurosci.* **10**, 272–282 (2009).
- Suh, J., Rivest, A. J., Nakashiba, T., Tominaga, T. & Tonegawa, S. Entorhinal cortex layer III input to the hippocampus is crucial for temporal association memory. *Science* **334**, 1415–1420 (2011).
- Nakashiba, T. et al. Young dentate granule cells mediate pattern separation, whereas old granule cells facilitate pattern completion. *Cell* **149**, 188–201 (2012).
- Nakashiba, T., Young, J. Z., McHugh, T. J., Buhl, D. L. & Tonegawa, S. Transgenic inhibition of synaptic transmission reveals role of CA3 output in hippocampal learning. *Science* **319**, 1260–1264 (2008).
- Tsien, J. Z., Huerta, P. T. & Tonegawa, S. The essential role of hippocampal CA1 NMDA receptor-dependent synaptic plasticity in spatial memory. *Cell* **87**, 1327–1338 (1996).
- Chevalayre, V. & Siegelbaum, S. A. Strong CA2 pyramidal neuron synapses define a powerful disinaptic cortico-hippocampal loop. *Neuron* **66**, 560–572 (2010).
- Lorente de Nó, R. Studies on the structure of the cerebral cortex. II. Continuation of the study of the ammonic system. *J. Psychol. Neurol.* **46**, 113–177 (1934).
- Franklin, K. & Paxinos, G. *The Mouse Brain in Stereotaxic Coordinates* (Academic, 2007).

10. Lein, E. S., Callaway, E. M., Albright, T. D. & Gage, F. H. Redefining the boundaries of the hippocampal CA2 subfield in the mouse using gene expression and 3-dimensional reconstruction. *J. Comp. Neurol.* **485**, 1–10 (2005).
11. Fanselow, M. S. & Dong, H. W. Are the dorsal and ventral hippocampus functionally distinct structures? *Neuron* **65**, 7–19 (2010).
12. Lee, S. E. *et al.* RGS14 is a natural suppressor of both synaptic plasticity in CA2 neurons and hippocampal-based learning and memory. *Proc. Natl Acad. Sci. USA* **107**, 16994–16998 (2010).
13. Cui, Z., Gerfen, C. R. & Young, W. S. III. Hypothalamic and other connections with dorsal CA2 area of the mouse hippocampus. *J. Comp. Neurol.* **521**, 1844–1866 (2013).
14. Kohara, K. *et al.* Cell type-specific genetic and optogenetic tools reveal hippocampal CA2 circuits. *Nature Neurosci.* 10.1038/nn.3614 (2013).
15. Wall, N. R., De La Parra, M., Callaway, E. M. & Kreitzer, A. C. Differential innervation of direct- and indirect-pathway striatal projection neurons. *Neuron* **79**, 347–360 (2013).
16. Rowland, D. C. *et al.* Transgenically targeted rabies virus demonstrates a major monosynaptic projection from hippocampal area CA2 to medial entorhinal layer II neurons. *J. Neurosci.* **33**, 14889–14898 (2013).
17. Boyden, E. S., Zhang, F., Bamberg, E., Nagel, G. & Deisseroth, K. Millisecond-timescale, genetically targeted optical control of neural activity. *Nature Neurosci.* **8**, 1263–1268 (2005).
18. Hargreaves, E. L., Rao, G., Lee, I. & Knierim, J. J. Major dissociation between medial and lateral entorhinal input to dorsal hippocampus. *Science* **308**, 1792–1794 (2005).
19. Hensler, J. G. Serotonergic modulation of the limbic system. *Neurosci. Biobehav. Rev.* **30**, 203–214 (2006).
20. Pan, W. X. & McNaughton, N. The supramammillary area: its organization, functions and relationship to the hippocampus. *Prog. Neurobiol.* **74**, 127–166 (2004).
21. Young, W. S., Li, J., Wersinger, S. R. & Palkovits, M. The vasopressin 1b receptor is prominent in the hippocampal area CA2 where it is unaffected by restraint stress or adrenalectomy. *Neuroscience* **143**, 1031–1039 (2006).
22. Wersinger, S. R., Ginns, E. I., O'Carroll, A. M., Lolait, S. J. & Young, W. S. III. Vasopressin V1b receptor knockout reduces aggressive behavior in male mice. *Mol. Psychiatry* **7**, 975–984 (2002).
23. DeVito, L. M. *et al.* Vasopressin 1b receptor knock-out impairs memory for temporal order. *J. Neurosci.* **29**, 2676–2683 (2009).
24. Stevenson, E. L. & Caldwell, H. K. The vasopressin 1b receptor and the neural regulation of social behavior. *Horm. Behav.* **61**, 277–282 (2012).
25. Kogan, J. H., Frankland, P. W. & Silva, A. J. Long-term memory underlying hippocampus-dependent social recognition in mice. *Hippocampus* **10**, 47–56 (2000).
26. Ferguson, J. N. *et al.* Social amnesia in mice lacking the oxytocin gene. *Nature Genet.* **25**, 284–288 (2000).
27. Brennan, P. A. & Zufall, F. Pheromonal communication in vertebrates. *Nature* **444**, 308–315 (2006).
28. Corkin, S. What's new with the amnesic patient H.M.? *Nature Rev. Neurosci.* **2**, 153–160 (2002).
29. Benes, F. M., Kwok, E. W., Vincent, S. L. & Todtenkopf, M. S. A reduction of nonpyramidal cells in sector CA2 of schizophrenics and manic depressives. *Biol. Psychiatry* **44**, 88–97 (1998).
30. Meyer-Lindenberg, A., Domes, G., Kirsch, P. & Heinrichs, M. Oxytocin and vasopressin in the human brain: social neuropeptides for translational medicine. *Nature Rev. Neurosci.* **12**, 524–538 (2011).

Acknowledgements We thank T. R. Reardon for providing the rabies virus; J. Kupferman and F. Lema for experimental assistance; and C. Denny, Z. Donaldson, R. Hen, J. Gordon, J. Basu and M. Russo for discussions and comments on the manuscript. This work was supported by a Ruth L. Kirschstein F30 National Research Service Award from the National Institute of Mental Health (F.L.H.) and the Howard Hughes Medical Institute (S.A.S.).

Author Contributions F.L.H. planned and performed the experiments, analysed the data and wrote the manuscript. S.A.S. oversaw the overall execution of the project, contributed to the experimental design and the interpretation of the data, provided financial support and helped write the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.A.S. (sas8@columbia.edu).

Mouse liver repopulation with hepatocytes generated from human fibroblasts

Saiyong Zhu¹, Milad Rezvani², Jack Harbell³, Aras N. Mattis^{2,4,5}, Alan R. Wolfe⁶, Leslie Z. Benet⁶, Holger Willenbring^{2,3,4} & Sheng Ding^{1,7}

Human induced pluripotent stem cells (iPSCs) have the capability of revolutionizing research and therapy of liver diseases by providing a source of hepatocytes for autologous cell therapy and disease modelling. However, despite progress in advancing the differentiation of iPSCs into hepatocytes (iPSC-Heps) *in vitro*^{1–3}, cells that replicate the ability of human primary adult hepatocytes (aHeps) to proliferate extensively *in vivo* have not been reported. This deficiency has hampered efforts to recreate human liver diseases in mice, and has cast doubt on the potential of iPSC-Heps for liver cell therapy. The reason is that extensive post-transplant expansion is needed to establish and sustain a therapeutically effective liver cell mass in patients, a lesson learned from clinical trials of aHep transplantation⁴. Here, as a solution to this problem, we report the generation of human fibroblast-derived hepatocytes that can repopulate mouse livers. Unlike current protocols for deriving hepatocytes from human fibroblasts, ours did not generate iPSCs but cut short reprogramming to pluripotency to generate an induced multipotent progenitor cell (iMPC) state from which endoderm progenitor cells and subsequently hepatocytes (iMPC-Heps) could be efficiently differentiated. For this purpose we identified small molecules that aided endoderm and hepatocyte differentiation without compromising proliferation. After transplantation into an immune-deficient mouse model of human liver failure, iMPC-Heps proliferated extensively and acquired levels of hepatocyte function similar to those of aHeps. Unfractionated iMPC-Heps did not form tumours, most probably because they never entered a pluripotent state. Our results establish the feasibility of significant liver repopulation of mice with human hepatocytes generated *in vitro*, which removes a long-standing roadblock on the path to autologous liver cell therapy.

Current protocols for directed hepatocyte differentiation of iPSCs (or human embryonic stem cells; ESCs) produce cells that express many functions of human primary adult hepatocytes (aHeps)^{1–3}. Some functions of iPSC/ESC-Heps, such as cytochrome P450 (CYP450) enzyme expression, are underdeveloped, whereas others, such as albumin (ALB) synthesis, are near normal. Human serum ALB (HSA) levels have therefore been used to follow the fate and measure the expansion of iPSC/ESC-Heps transplanted into mice. Since the advent of protocols for hepatocyte differentiation of iPSCs/ESCs, numerous results from transplantation of these cells into mice have been reported. All of these studies, even those in which transplanted iPSC/ESC-Heps had a growth advantage, failed to reach HSA levels above 2 µg ml⁻¹ (refs 5–7), which is more than 1,000-fold lower than what has been achieved with aHeps⁸ and in our experience reflects less than 0.05% liver repopulation (data not shown). These disappointing results have been largely attributed to failure of iPSC/ESC-Heps to proliferate after transplantation.

Reprogramming of somatic cells to a stable pluripotent state followed by differentiation into another cell type is a complicated process.

Alternatively, the developmentally plastic state established soon after overexpression of the reprogramming factors can be harnessed for lineage conversion. Using this approach we and others previously induced human fibroblasts to assume a multipotent progenitor cell (iMPC) state that allowed efficient differentiation into myeloid⁹ or smooth muscle and endothelial cells^{10,11}. Because these iMPC derivatives could proliferate extensively, we reasoned that this method could be used to generate hepatocytes that were not compromised by growth arrest.

To investigate this possibility we transduced 10⁴ human fibroblasts with retroviruses expressing OCT4, SOX2 and KLF4 (ref. 12) and replated them for reprogramming into endoderm in medium containing established growth factors and the small molecule CHIR99021 (CHIR)^{1–3} (Fig. 1). Only 14 days later, we detected expression of the endoderm-specific genes *SOX17* and *FOXA2* by quantitative PCR with reverse transcription (qRT-PCR) (Extended Data Fig. 1a). While only about 20 *SOX17*-positive and *FOXA2*-positive colonies formed under these

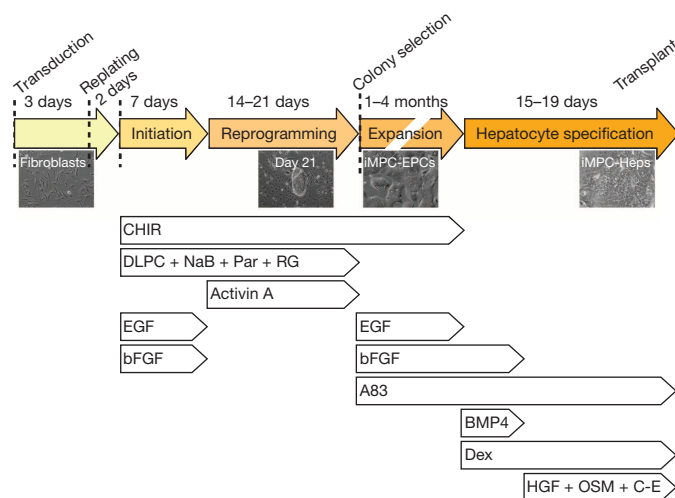


Figure 1 | Protocol for stepwise iMPC-Hep generation. Reprogramming of human fibroblasts to endoderm was initiated in medium containing CHIR (glycogen synthase kinase 3β inhibitor), dilauroyl phosphatidylcholine (DLPC; liver receptor homologue-1 agonist), the epigenetic modifiers sodium butyrate (NaB; histone deacetylase inhibitor), Parnate (Par; lysine-specific demethylase 1 inhibitor) and RG108 (RG; DNA methyltransferase inhibitor), and epidermal growth factor (EGF) and basic fibroblast growth factor (bFGF). To promote reprogramming, EGF and bFGF were replaced with activin A. Individual iMPC-EPC colonies were expanded in medium containing CHIR, EGF, bFGF and A83 (transforming growth factor-β type I receptor inhibitor). For hepatocyte specification, medium containing bFGF, A83, bone morphogenetic protein 4 (BMP4), dexamethasone (Dex), hepatocyte growth factor (HGF), oncostatin M (OSM) and the Notch inhibitor Compound E (C-E) was used.

¹Gladstone Institute of Cardiovascular Disease, 1650 Owens Street, San Francisco, California 94158, USA. ²Eli and Edythe Broad Center of Regeneration Medicine and Stem Cell Research, University of California San Francisco, 35 Medical Center Way, San Francisco, California 94143, USA. ³Department of Surgery, Division of Transplantation, University of California San Francisco, 505 Parnassus Avenue, San Francisco, California 94143, USA. ⁴Liver Center, University of California San Francisco, 1001 Potrero Avenue, San Francisco, California 94110, USA. ⁵Department of Pathology, University of California San Francisco, 505 Parnassus Avenue, San Francisco, California 94143, USA. ⁶Department of Bioengineering and Therapeutic Sciences, University of California San Francisco, 533 Parnassus Avenue, San Francisco, California 94143, USA. ⁷Department of Pharmaceutical Chemistry, University of California San Francisco, 600 16th Street, San Francisco, California 94158, USA.

conditions, exposing the cells to additional small molecules known to promote reprogramming^{13,14} increased the number of colonies to more than 80 (Extended Data Fig. 1b–d and Supplementary Table 1).

Next we investigated whether *de novo* endoderm differentiation was preceded by a pluripotent state. We found no expression of the pluripotency-specific genes *OCT4* and *NANOG* even at the earliest stages of the reprogramming process (Extended Data Fig. 1e). Because avoiding a pluripotent state decreases the cells' tumour risk, we confirmed this result by TRA-1-60 (ref. 15) flow cytometry at the end of the reprogramming process (Extended Data Fig. 1f, g). In addition we monitored cultures undergoing reprogramming for FOXA2-positive cells, referred to as iMPC-derived endoderm progenitor cells (iMPC-EPCs), and NANOG-positive cells by immunostaining and flow cytometry (Extended Data Fig. 2a). We found FOXA2-positive cells as early as 16 days after initiating reprogramming, whereas NANOG-positive cells were always absent (Extended Data Fig. 2b–d). We also used doxycycline-inducible lentiviruses expressing *OCT4*, *SOX2* and *KLF4* to compare the dynamics of reprogramming to endoderm versus pluripotency (Extended Data Fig. 3a). We detected iMPC-EPC colonies in transduced cultures grown under iMPC-EPC reprogramming conditions for 21 days after only 7 days of treatment with doxycycline. In contrast, generation of iPSCs required treating the cultures with doxycycline for 14 days and growing them under iPSC reprogramming conditions for 30 days (Extended Data Fig. 3b). Our findings that human fibroblasts reprogram into iMPC-EPCs faster than into iPSCs and without expressing pluripotency markers show that our protocol does not produce a pluripotent intermediate stage, which confirms previous results from shortcutting reprogramming to pluripotency for lineage conversion^{9–11}.

We also determined whether iMPC-EPCs could be expanded *in vitro*, a prerequisite for producing the large quantities needed for human liver cell therapy. We found that combining CHIR with A83-01 (A83)¹⁶ increased iMPC-EPC colony size (Fig. 1 and Extended Data Fig. 4a). The addition of epidermal growth factor and basic fibroblast growth factor caused further expansion and facilitated passaging for more than

25 times, producing more than 10^{16} iMPC-EPCs from a single colony (Fig. 2a, b, Extended Data Fig. 4b and Supplementary Table 1). These cells showed high viability after cryopreservation (data not shown).

Expanding iMPC-EPCs maintained endoderm differentiation, as demonstrated by positive immunostaining for FOXA2 and SOX17 and negative immunostaining for NANOG (Fig. 2a and Extended Data Fig. 4c). Expanding iMPC-EPCs acquired hepatocyte nuclear factor 4 α (HNF4 α) expression (Fig. 2a and Extended Data Fig. 4d), suggesting further specification. To define their stage of differentiation, we compared iMPC-EPCs with ESC-derived definitive endoderm cells (DECs) and primitive gut-tube endoderm cells (GECs)¹⁷ (Fig. 2c). We found that iMPC-EPCs resembled ESC-GECs, except for a lack of *OCT4* and *NANOG* gene expression. iMPC-EPCs also lacked expression of the ectoderm-specific and mesoderm-specific genes *PAX6* and *BRY*, suggesting commitment to endoderm differentiation. Further analyses showed that iMPC-EPCs had a propensity for differentiating into liver and pancreas but not into lung or intestine (Extended Data Fig. 4e, f and data not shown). These results establish the feasibility of using human fibroblasts to generate endoderm cells that share many characteristics with previously reported ESC/iPSC-derived endodermal progenitor cell lines¹⁸, but seem more lineage restricted and never entered a pluripotent state.

To further differentiate iMPC-EPCs into iMPC-Heps, we cultured them in medium containing factors that have been reported to drive the hepatic differentiation of iPSC-DECs^{1,3} (Fig. 1). These factors were effective in inducing expression of the fetal hepatocyte marker α -fetoprotein, but few cells expressed the more mature markers ALB and α_1 -antitrypsin (AAT) (Extended Data Fig. 5a). To improve hepatocyte differentiation, we screened small molecules for inducers of *ALB* gene expression, of which A83 and the Notch inhibitor Compound E¹⁶ were effective (Extended Data Fig. 5b). Because transforming growth factor- β and Notch signalling direct bipotential embryonic liver progenitor cells towards a biliary fate^{19,20}, our results suggest that inhibiting biliary differentiation promotes hepatocyte differentiation.

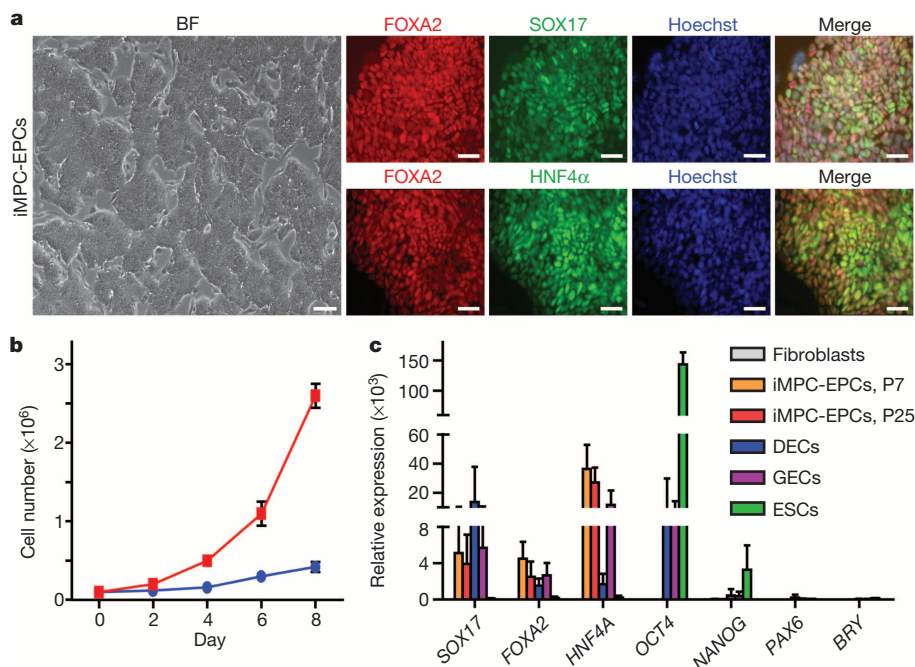


Figure 2 | Characterization of iMPC-EPCs. **a**, Bright-field (BF) microscopy shows the morphology of iMPC-EPCs at passage 25; immunostainings show expression of FOXA2, SOX17 and HNF4 α . Scale bars, 100 μ m. **b**, Expansion capacity of iMPC-EPCs (red) and parental fibroblasts (blue). Cell numbers were counted at the indicated time points. Results are means \pm s.e.m. for biological replicates ($n = 3$). **c**, qRT-PCR of genes specific for endoderm,

pluripotency, ectoderm or mesoderm in iMPC-EPCs (P7, passage 7; P25, passage 25) in comparison with parental fibroblasts, ESCs and ESC-derived DECs or GECs. Gene expression levels are shown relative to the fibroblasts. Results are means and s.e.m. for technical replicates ($n = 3$). All results were replicated in at least three independent experiments.

Like aHeps, iMPC-Heps were polygonal, were occasionally binucleate, and expressed the hepatocyte markers HNF4 α , ALB, AAT and cytokeratin 18 (Fig. 3a). iMPC-Heps also showed hepatocyte functions such as glycogen storage, lipid uptake and storage, and urea production (Extended Data Fig. 6a, b). Gene expression analysis showed that iMPC-Heps generally resembled human primary fetal hepatocytes (fHeps) (Fig. 3b), although some cells were less differentiated (Fig. 3c and Supplementary Table 1). Analysis of ALB secretion and CYP450 activities confirmed that iMPC-Heps were less mature than aHeps, but also showed that iMPC-Heps were more differentiated than iPSC-Heps generated as reported previously^{1,3} (Fig. 3d, e and Extended Data Fig. 6c). The medium used for iMPC-EPC/Hep generation did

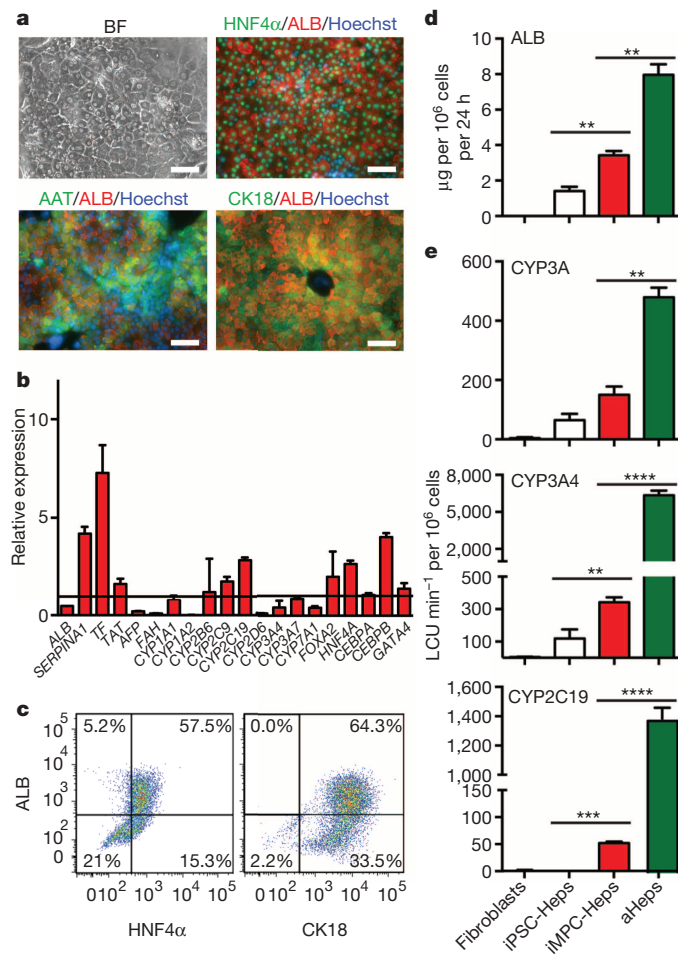
not produce iPSC-Heps with improved function, which underscores the importance of reprogramming-induced developmental plasticity in this process (Extended Data Fig. 6d).

To test whether iMPC-Heps can expand after transplantation, we transplanted 10^6 cells into FRG mice, an immune-deficient mouse model of human tyrosinaemia type I (ref. 8). The liver injury caused by this disease creates a growth advantage for differentiated hepatocytes but not for immature liver progenitor cells. Liver repopulation of FRG mice therefore requires both mature hepatocyte function and the ability to proliferate. To detect expansion of the transplanted iMPC-Heps, we measured HSA levels monthly for more than nine months. The earliest we could detect HSA was two months after transplantation (Fig. 4a), when levels were at most 140 ng ml^{-1} , but they increased continuously, reaching levels of up to $104 \mu\text{g ml}^{-1}$ six months later. By this time, HSA levels were tenfold higher in control FRG mice transplanted with 10^6 aHeps. The delayed onset but parallel upward trend of HSA levels in iMPC-Hep-transplanted mice, in comparison with control mice, suggested that iMPC-Heps were inferior to aHeps in engraftment efficiency and the need for post-transplant maturation, but not in the ability to proliferate. Indeed, we found that although iMPC-Heps generated significantly fewer repopulating nodules than aHeps (data not shown), these nodules grew markedly between three and nine months after transplantation (Extended Data Fig. 7a). Moreover, iMPC-Heps still proliferated nine months after transplantation (Fig. 4b). So far we have observed a maximum nodule size of 4,000 iMPC-Heps and a liver repopulation level of 2% (Extended Data Fig. 7b, c).

To determine whether iMPC-Heps matured after transplantation, we compared the global gene expression profiles of transplanted iMPC-Heps and aHeps. For this purpose we isolated nodules of iMPC-Heps and aHeps by laser-capture microscopy (Extended Data Fig. 7d, e) and analysed their RNA with microarrays. We found that iMPC-Heps and aHeps clustered closely together—very few genes were differentially expressed, none of which were of known importance for hepatocyte function (Fig. 4c, Extended Data Fig. 8a–e and Supplementary Table 2). We also compared cultured iMPC-Heps and freshly isolated aHeps, and found marked differences in gene expression between these two cell types, which illustrates the extensive maturation that iMPC-Heps underwent after transplantation. In fact, before transplantation, the gene expression profile of iMPC-Heps resembled that of iPSC-Heps. We confirmed the microarray results by qRT-PCR and immunostaining (Fig. 4d and Extended Data Fig. 9a–c). In addition we determined whether maturation of gene expression translated into normal function by measuring debrisoquine hydroxylation—a unique function of human hepatocytes executed by CYP2D6 (ref. 21)—in mice repopulated to similar levels with iMPC-Heps or aHeps. We found no difference in plasma 4-hydroxy-debrisoquine levels between these mice, which shows that CYP2D6 underwent maturation in iMPC-Heps from negligible expression levels *in vitro* to normal activity *in vivo* (Figs 3b and 4e). In accord with a need for post-transplant maturation of iMPC-Heps, which is reminiscent of iPSC/ESC-derived pancreatic β -cells²², we found that iMPC-Hep transplantation improved the survival of mice with chronic liver failure, but not that of those with acute liver failure (Extended Data Fig. 10a, b).

Finally, we ruled out fusion with mouse hepatocytes as the reason for post-transplant maturation and proliferation of iMPC-Heps (Fig. 4f), and investigated the origin of dysplastic nodules observed in some iMPC-Hep and aHep recipients (Extended Data Fig. 10c). Absence of differentiation-independent, human-specific β_2 -microglobulin expression showed that these nodules originated from mouse cells, thus representing a known complication of tyrosinaemia type I (Extended Data Fig. 10d)²³.

Our results show that iMPC-Heps are not affected by limitations of iPSC/ESC-Heps generated with current protocols, particularly deficiencies in *in vivo* efficacy and safety. Although many aspects of iMPC-Hep generation and transplantation remain to be explored and improved, the fact that these cells can fully mature and proliferate for months after



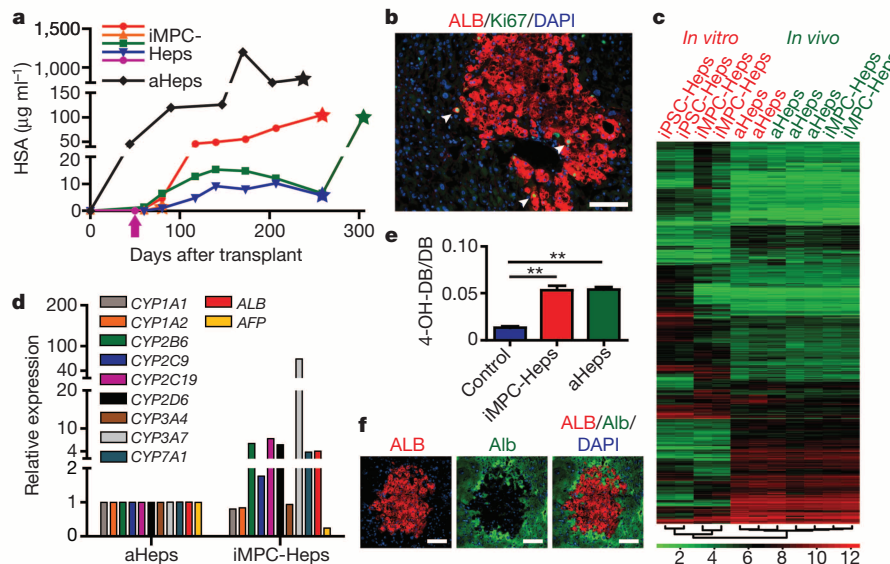


Figure 4 | Post-transplant proliferation and maturation of iMPC-Heps. **a**, HSA levels in recipients of iMPC-Heps or aHeps. Stars indicate time points of analysis. Arrow marks fatality. **b**, Co-immunostaining for human-specific ALB and Ki67 identifies proliferating iMPC-Heps (arrowheads) in the periphery of a repopulating nodule. Scale bar, 100 µm. **c**, Heatmap of 1,299 genes differentially expressed between iMPC-Heps, freshly isolated aHeps and iMPC-Heps before (*in vitro*) and after (*in vivo*) transplantation. Multiple nodules were pooled to generate a sample. Genes with expression levels below background (\log_2 normalized expression < 3) and genes not varying over all samples (s.d. expression < 1) were filtered out. Hierarchical clustering was performed with the *hclust* function in R v.2.15.1. **d**, qRT-PCR of the samples used for microarray analysis shows mean hepatocyte marker gene expression in

iMPC-Heps relative to aHeps *in vivo*. **e**, Analysis of human-specific CYP2D6-mediated debrisoquine (DB) metabolism in iMPC-Hep- or aHep-repopulated mice by liquid chromatography–tandem mass spectrometry (LC–MS/MS). Plasma levels of DB and its metabolite 4-hydroxy-debrisoquine (4-OH-DB) peaked 1 h after gavage. Molar 4-OH-DB/DB ratios at 1 h are shown, calculated as the mean of the ratios for repeat injections ($n = 3$). Error bars represent analytical s.e.m.; Student's *t*-test, two asterisks, $P < 0.01$. **f**, Co-immunostaining with human-specific ALB and mouse-specific Alb antibodies shows an absence of double-positive cells, which excludes fusion of iMPC-Heps with mouse hepatocytes. Scale bars, 100 µm. All results (except microarray analysis) were replicated in at least two independent experiments.

transplantation establishes them as promising candidates for *in vivo* modelling and autologous therapy of human liver diseases.

METHODS SUMMARY

Generation of iMPC-Heps from human fibroblasts. After retroviral transduction, human fibroblasts were grown in reprogramming initiation medium for 7 days, followed by culture in endoderm differentiation medium for 14–21 days. iMPC-EPC colonies were selected at day 21–28 and grown on mouse embryonic fibroblasts in endoderm expansion medium for 1–4 months. For differentiation into iMPC-Heps, iMPC-EPCs were first expanded for 4–5 days, followed by culture in hepatocyte differentiation medium for 4 days and then in hepatocyte maturation medium for 7–10 days. Media were supplemented with the growth factors and small molecules indicated in Fig. 1.

iMPC-Hep transplantation and analysis. iMPC-Heps were injected intrasplenically into FRG mice preconditioned by withdrawal of the drug 2-(2-nitro-4-fluoromethylbenzoyl)-1,3-cyclohexanedione (NTBC) and retro-orbital injection of an adenovirus expressing urokinase plasminogen activator. Mice underwent repeated cycles of NTBC: absent for 7–10 days and present for 2–3 days. Liver repopulation was monitored by HSA enzyme-linked immunosorbent assay (ELISA).

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 12 June 2013; accepted 10 January 2014.

Published online 23 February 2014.

- Si-Tayeb, K. *et al.* Highly efficient generation of human hepatocyte-like cells from induced pluripotent stem cells. *Hepatology* **51**, 297–305 (2010).
- Rashid, S. T. *et al.* Modeling inherited metabolic disorders of the liver using human induced pluripotent stem cells. *J. Clin. Invest.* **120**, 3127–3136 (2010).
- Ma, X. *et al.* Highly efficient differentiation of functional hepatocytes from human induced pluripotent stem cells. *Stem Cells Transl. Med.* **2**, 409–419 (2013).
- Puppi, J. *et al.* Improving the techniques for human hepatocyte transplantation: report from a consensus meeting in London. *Cell Transplant.* **21**, 1–10 (2012).
- Liu, H., Kim, Y., Sharkis, S., Marchionni, L. & Jang, Y. Y. *In vivo* liver regeneration potential of human induced pluripotent stem cells from diverse origins. *Sci. Transl. Med.* **3**, 82ra39 (2011).

- Basma, H. *et al.* Differentiation and transplantation of human embryonic stem cell-derived hepatocytes. *Gastroenterology* **136**, 990–999 (2009).
- Woo, D. H. *et al.* Direct and indirect contribution of human embryonic stem cell-derived hepatocyte-like cells to liver repair in mice. *Gastroenterology* **142**, 602–611 (2012).
- Azuma, H. *et al.* Robust expansion of human hepatocytes in *Fah^{-/-}/Rag2^{-/-}/Il2rg^{-/-}* mice. *Nature Biotechnol.* **25**, 903–910 (2007).
- Szabo, E. *et al.* Direct conversion of human fibroblasts to multilineage blood progenitors. *Nature* **468**, 521–526 (2010).
- Kurian, L. *et al.* Conversion of human fibroblasts to angioblast-like progenitor cells. *Nature Methods* **10**, 77–83 (2013).
- Li, J. *et al.* Conversion of human fibroblasts to functional endothelial cells by defined factors. *Arterioscler. Thromb. Vasc. Biol.* **33**, 1366–1375 (2013).
- Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872 (2007).
- Zhu, S. *et al.* Reprogramming of human primary somatic cells by OCT4 and chemical compounds. *Cell Stem Cell* **7**, 651–655 (2010).
- Lee, J. M. *et al.* A nuclear-receptor-dependent phosphatidylcholine pathway with antidiabetic effects. *Nature* **474**, 506–510 (2011).
- Chan, E. M. *et al.* Live cell imaging distinguishes bona fide human iPS cells from partially reprogrammed cells. *Nature Biotechnol.* **27**, 1033–1037 (2009).
- Li, W. *et al.* Rapid induction and long-term self-renewal of primitive neural precursors from human embryonic stem cells by small molecule inhibitors. *Proc. Natl Acad. Sci. USA* **108**, 8299–8304 (2011).
- Wang, P., Rodriguez, R. T., Wang, J., Ghodasara, A. & Kim, S. K. Targeting SOX17 in human embryonic stem cells creates unique strategies for isolating and analyzing developing endoderm. *Cell Stem Cell* **8**, 335–346 (2011).
- Cheng, X. *et al.* Self-renewing endodermal progenitor lines generated from human pluripotent stem cells. *Cell Stem Cell* **10**, 371–384 (2012).
- Clotman, F. *et al.* Control of liver cell fate decision by a gradient of TGF β signaling modulated by Onecut transcription factors. *Genes Dev.* **19**, 1849–1854 (2005).
- Kodama, Y., Hijikata, M., Kageyama, R., Shimotohno, K. & Chiba, T. The role of notch signaling in the development of intrahepatic bile ducts. *Gastroenterology* **127**, 1775–1786 (2004).
- Chen, A. A. *et al.* Humanized mice with ectopic artificial liver tissues. *Proc. Natl Acad. Sci. USA* **108**, 11842–11847 (2011).
- Kroon, E. *et al.* Pancreatic endoderm derived from human embryonic stem cells generates glucose-responsive insulin-secreting cells *in vivo*. *Nature Biotechnol.* **26**, 443–452 (2008).
- Willenbring, H. *et al.* Loss of p21 permits carcinogenesis from chronically damaged liver and kidney epithelial cells despite unchecked apoptosis. *Cancer Cell* **14**, 59–67 (2008).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the Gladstone Institutes' Bioinformatics Core for data analysis, A. Grimm for discussion, and P. Derish for manuscript editing. H.W. is supported by funding from the California Institute for Regenerative Medicine (CIRM; RN2-00950) and the National Institutes of Health (NIH; P30 DK26743). S.D. is supported by funding from CIRM, NIH and the Gladstone Institutes. S.Z. is supported by CIRM research training grant TG2-01160. M.R. is a research fellow in the Biomedical Exchange Program funded by the German Academic Exchange Service. J.H. is an Ethicon-Society of University Surgeons Fellow. A.N.M. is supported by CIRM research training grant TG2-01153.

Author Contributions S.Z., M.R. and J.H. are joint first authors; H.W. and S.D. are joint senior authors. S.Z., M.R., J.H., H.W. and S.D. designed the experiments. S.Z. and A.N.M.

performed the reprogramming and directed differentiation experiments. M.R. and J.H. performed the transplantation experiments. M.R., J.H. and A.N.M. analysed the transplantation experiments and performed additional *in vitro* analyses. A.R.W. and L.Z.B. performed the liquid chromatography–tandem mass spectrometry analyses. S.Z., M.R., J.H., H.W. and S.D. wrote the manuscript. H.W. and S.D. edited the manuscript. All authors read and approved the final manuscript.

Author Information Results of the microarray analysis have been deposited in the Gene Expression Omnibus database under accession no. GSE52309. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.W. (willenbringh@stemcell.ucsf.edu) or S.D. (sheng.ding@gladstone.ucsf.edu).

Constitutional and somatic rearrangement of chromosome 21 in acute lymphoblastic leukaemia

Yilong Li^{1*}, Claire Schwab^{2*}, Sarra L. Ryan^{2*}, Elli Papaemmanuil¹, Hazel M. Robinson³, Patricia Jacobs⁴, Anthony V. Moorman², Sara Dyer^{3,5}, Julian Borrow^{3,5}, Mike Griffiths^{3,5}, Nyla A. Heerema⁶, Andrew J. Carroll⁷, Polly Talley⁸, Nick Bown⁹, Nick Telford¹⁰, Fiona M. Ross⁴, Lorraine Gaunt¹¹, Richard J. Q. McNally¹², Bryan D. Young², Paul Sinclair², Vikki Rand², Manuel R. Teixeira^{13,14}, Olivia Joseph¹, Ben Robinson¹, Mark Maddison¹, Nicole Dastugue¹⁵, Peter Vandenberghe^{16,17}, Claudia Haeflrich¹⁸, Philip J. Stephens¹, Jiqiu Cheng^{16,17,19}, Peter Van Loo^{1,16,17}, Michael R. Stratton¹, Peter J. Campbell^{1,20*} & Christine J. Harrison^{2*}

Changes in gene dosage are a major driver of cancer, known to be caused by a finite, but increasingly well annotated, repertoire of mutational mechanisms¹. This can potentially generate correlated copy-number alterations across hundreds of linked genes, as exemplified by the 2% of childhood acute lymphoblastic leukaemia (ALL) with recurrent amplification of megabase regions of chromosome 21 (iAMP21)^{2,3}. We used genomic, cytogenetic and transcriptional analysis, coupled with novel bioinformatic approaches, to reconstruct the evolution of iAMP21 ALL. Here we show that individuals born with the rare constitutional Robertsonian translocation between chromosomes 15 and 21, rob(15;21)(q10;q10)c, have approximately 2,700-fold increased risk of developing iAMP21 ALL compared to the general population. In such cases, amplification is initiated by a chromothripsis event involving both sister chromatids of the Robertsonian chromosome, a novel mechanism for cancer predisposition. In sporadic iAMP21, breakage-fusion-bridge cycles are typically the initiating event, often followed by chromothripsis. In both sporadic and rob(15;21)c-associated iAMP21, the final stages frequently involve duplications of the entire abnormal chromosome. The end-product is a derivative of chromosome 21 or the rob(15;21)c chromosome with gene dosage optimized for leukaemic potential, showing constrained copy-number levels over multiple linked genes. Thus, dicentric chromosomes may be an important precipitant of chromothripsis, as we show rob(15;21)c to be constitutionally dicentric and breakage-fusion-bridge cycles generate dicentric chromosomes somatically. Furthermore, our data illustrate that several cancer-specific mutational processes, applied sequentially, can coordinate to fashion copy-number profiles over large genomic scales, incrementally refining the fitness benefits of aggregated gene dosage changes.

Acute lymphoblastic leukaemia (ALL) is the most common childhood cancer, with an annual incidence of 35/million children aged 0–14 years⁴. Approximately 2% of these cases show intrachromosomal amplification of one copy of chromosome 21, iAMP21, which defines a distinct ALL subgroup^{3,5} with prognostic and therapeutic implications^{6,7}.

In 95 patients with iAMP21 ALL enrolled in UK clinical trials, we found 3 (3.2%) with a constitutional Robertsonian translocation between chromosomes 15 and 21, rob(15;21)(q10;q10)c. In other ALL trials, we identified a further nine cases of iAMP21 associated with rob(15;21)c, at a similar fraction of cases. Robertsonian translocations are rearrangements between the short arms of acrocentric chromosomes (namely,

13–15, 21–22). They are found in about 1 in 1,000 newborns^{8,9}, but rob(15;21)c accounts for only 0.5–1% of these. To confirm this, we interrogated cytogenetics databases. Only three patients among 93,000 referrals for haematological malignancies to the Munich Leukaemia Laboratory and West Midlands Regional Genetics Laboratory carried rob(15;21)c. Similarly, only 16 cases were found among approximately 300,000 referrals to UK regional cytogenetics laboratories for investigation of infertility or previous Down syndrome birth.

From these data, we estimate the risk of iAMP21 ALL in carriers of rob(15;21)c to be increased approximately 2,700-fold over the general population (Supplementary Table 1). This association is remarkably specific. All patients in this study with rob(15;21)c had iAMP21 ALL, implying that they are not predisposed to other forms of ALL, nor other cancers, as far as we can ascertain. Furthermore, the only Robertsonian translocation associated with iAMP21 ALL was rob(15;21)c. For clarity, we use ‘rob(15;21)c’ to denote the germline configuration and ‘der(15;21)’ to describe the rearranged and amplified chromosome in these cases.

Using cytogenetics, fluorescence *in situ* hybridization (FISH) and copy-number profiling, we studied 21 patients with sporadic iAMP21 ALL and 12 patients with ALL associated with rob(15;21)c. Five sporadic iAMP21 and four cases associated with rob(15;21)c were sequenced to identify genomic rearrangements¹⁰ (Supplementary Table 2; Extended Data Fig. 1; Supplementary Fig. 1). We applied deductive approaches, supported by confirmatory simulations, to reconstruct principles underlying the temporal evolution of iAMP21 ALL. This reasoning is explored in considerable detail, together with a sample-by-sample analysis, in Supplementary Results, Extended Data Figs 3–8, Supplementary Tables 3–6 and Supplementary Figs 4–23.

The broad themes are illustrated by two representative cases (Figs 1, 2). In PD9020a, a patient with sporadic iAMP21, the boundaries of the amplified region are demarcated by fold-back inversion rearrangements (Fig. 1a). These indicate breakage-fusion-bridge (BFB) repair¹¹, previously proposed to trigger iAMP21^{2,12,13}. Breakage-fusion-bridge repair is a mutational process initiated by a telomeric double strand (ds) DNA break that is replicated in S phase. In G2, the two copies of the dsDNA break are fused by non-homologous end-joining (marked (1) in Fig. 1a). This creates a dicentric chromosome in which the two centromeres are pulled to opposite poles during mitosis, forming an anaphase bridge. With cytokinesis, the bridge breaks, and the process can repeat in the next cell cycle (marked (2) in Fig. 1a). In the region

¹Cancer Genome Project, Wellcome Trust Sanger Institute, Hinxton CB10 1SA, UK. ²Leukaemia Research Cytogenetics Group, Northern Institute for Cancer Research, Newcastle University, Newcastle upon Tyne NE2 4HH, UK. ³West Midlands Regional Genetics Laboratory, Birmingham Women's NHS Foundation Trust, Birmingham B15 2TG, UK. ⁴Wessex Regional Genetics Laboratory, Salisbury NHS Foundation Trust, Salisbury SP2 8BJ, UK. ⁵School of Cancer Sciences, University of Birmingham, Birmingham B15 2TT, UK. ⁶Department of Pathology, The Ohio State University, Columbus, Ohio 43210, USA. ⁷Department of Genetics, University of Alabama at Birmingham, Birmingham, Alabama 35233, USA. ⁸Sheffield Diagnostic Genetics Service, Sheffield Children's NHS Foundation Trust, Sheffield S10 2TH, UK. ⁹Cytogenetics Laboratory, Northern Genetics Service, Newcastle upon Tyne NE7 7DN, UK. ¹⁰Oncology Cytogenetics, The Christie NHS Foundation Trust, Manchester M20 4BX, UK. ¹¹Regional Cytogenetics Unit, Genetic Medicine, Central Manchester University Hospitals NHS Foundation Trust, Saint Mary's Hospital, Manchester M13 9WL, UK. ¹²Institute of Health and Society, Newcastle University, Newcastle upon Tyne NE2 4AX, UK. ¹³Genetics Department, Portuguese Oncology Institute, Porto University, 4200-072 Porto, Portugal. ¹⁴Biomedical Sciences Institute (ICBAS), Porto University, 4200-072 Porto, Portugal. ¹⁵Laboratoire d'Hématologie, Génétique des Hémapathies, Hôpital Purpan, 31059 Toulouse, France. ¹⁶Center for Human Genetics, University Hospital Leuven, 3000 Leuven, Belgium. ¹⁷KU Leuven, 3000 Leuven, Belgium. ¹⁸MLL Munich Leukemia Laboratory, Munich 81377, Germany. ¹⁹Department of Electrical Engineering — ESAT, University of Leuven, 3000 Leuven, Belgium. ²⁰Department of Haematology, University of Cambridge, Cambridge CB2 2XY, UK.

*These authors contributed equally to this work.

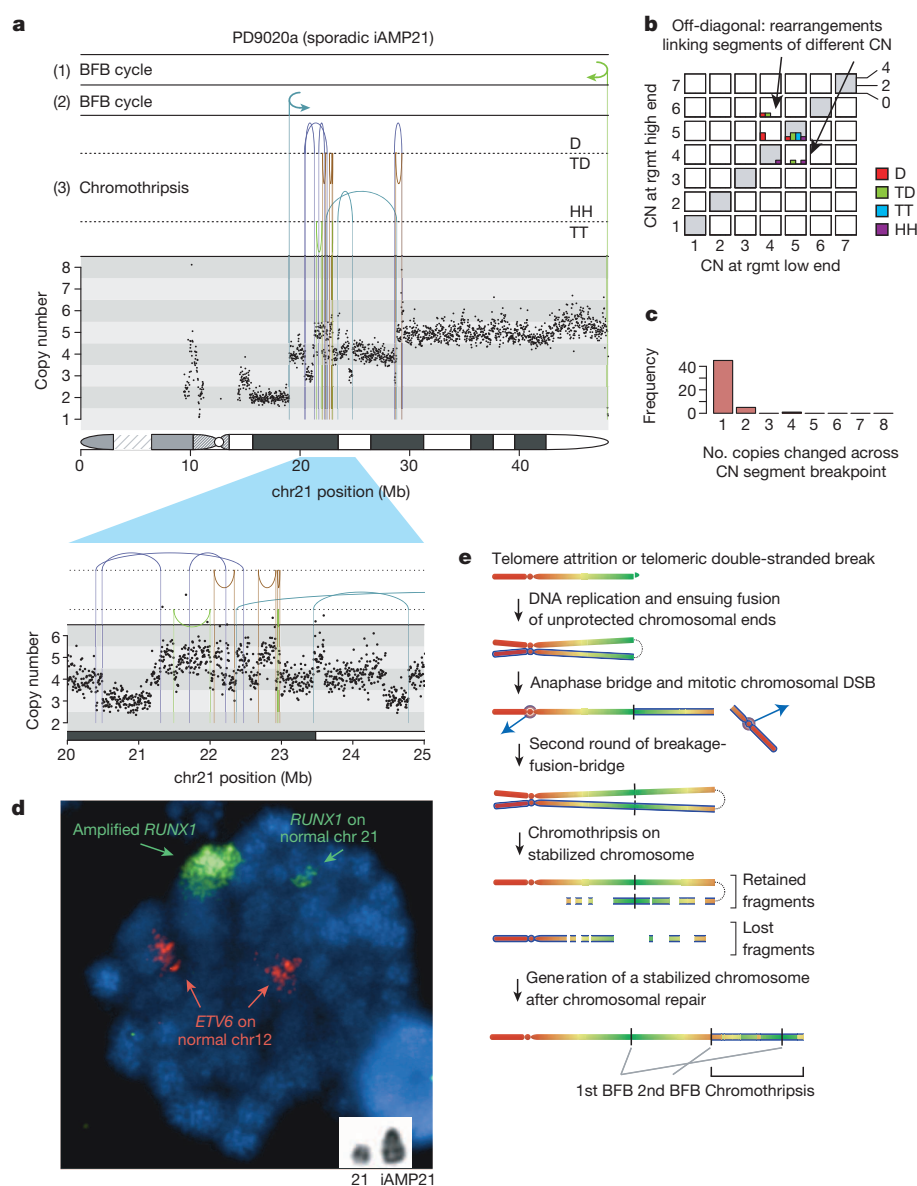


Figure 1 | Rearrangements of chromosome 21 in patient PD9020a. **a**, Rearrangement and copy-number pattern. The temporal order of the three major rearrangement events are marked (1), (2) and (3). Rearrangements are separated on the basis of their orientation: D, deletion-type; TD, tandem duplication-type; HH, head-to-head inverted; TT, tail-to-tail inverted. **b**, Copy-number (CN) jump distribution, showing the copy number at each end of each rearrangement (rgmt). **c**, Copy-number step distribution, showing the distribution in magnitude of copy-number change at copy-number segmentation breakpoints. **d**, Metaphase showing multiple signals for *RUNX1* clustered on a single chromosome (large green signal) compared to the normal chromosome 21 (small paired green signals). The red signals indicate two normal copies of *ETV6* on the chromosomes 12. Inset shows a partial G-banded karyotype of chromosomes 21. **e**, Model for the evolution of the iAMP21 chromosome. At each stage, newly synthesized sister chromatids are distinguished by a blue outline.

between the fold-back inversions in PD9020a, we also found a cluster of back-and-forth rearrangements (marked (3)) of all four possible orientations, associated with copy-number profiles that oscillate among three states (Fig. 1a, zoomed-in panel). These clusters bear the hallmarks of chromothripsis¹⁴ (Extended data Table 1; Extended Data Fig. 2; Supplementary Results), a mutational process in which a one-off catastrophic event shatters one or a few chromosomal regions leading to large numbers of localized genomic rearrangements^{15,16}.

Two features of this genomic architecture allow reconstruction of the temporal evolution of the iAMP21 chromosome. First, the rearrangements frequently link together genomic segments of different copy number (off-diagonal histograms, Fig. 1b). Second, as we traverse chromosome 21 from first to last base-pair, the copy-number segments on either side of each breakpoint position typically differ in copy number by one (Fig. 1c). FISH confirms widespread *RUNX1* signals along the iAMP21 chromosome (Fig. 1d). On the basis of reasoning outlined in detail in Supplementary Results, these features indicate that chromothripsis occurred after two BFB cycles and was likely the final major event, stabilizing the chromosome (Fig. 1e).

In PD7170a, a der(15;21) iAMP21 derived from rob(15;21)c, the picture is dominated by a series of back-and-forth rearrangements spanning chromosomes 15 and 21 (Fig. 2a–c). Cytogenetic and FISH studies

confirmed that it was the Robertsonian chromosome undergoing rearrangement (Fig. 2d–e). A sizable number of rearrangements link together segments of different copy number (off-diagonal histograms, Fig. 2b), and copy number oscillates among three, rather than two, states. Together with occasional inverted rearrangements with no breakpoints between the two joined ends, this pattern indicates that chromothripsis was the initiating event, and that the chromothripsis process involved both sister chromatids of the Robertsonian chromosome (Extended Data Figs 7, 8; Supplementary results; Supplementary Table 4). Importantly, the shattered sister chromatids are repaired into one derivative chromosome, thereby amplifying the copy number of some chromosomal regions (Fig. 2f). The amplification was completed by whole-chromosome duplication of the der(15;21) chromosome through isochromosome formation (Fig. 2d).

These broad temporal sequences of events are reflected in the other samples (Fig. 3; Extended Data Fig. 6; Supplementary Results). In the other four sequenced cases of sporadic iAMP21, a telomeric fold-back inversion suggests at least one BFB cycle. In each, this was a critical early event, defining the break between the most amplified region of chromosome 21 and subtelomeric loss. Chromothripsis occurred after the BFB cycles in three cases. Finally, partial or whole-chromosome duplications usually completed the evolution. In the other three sequenced

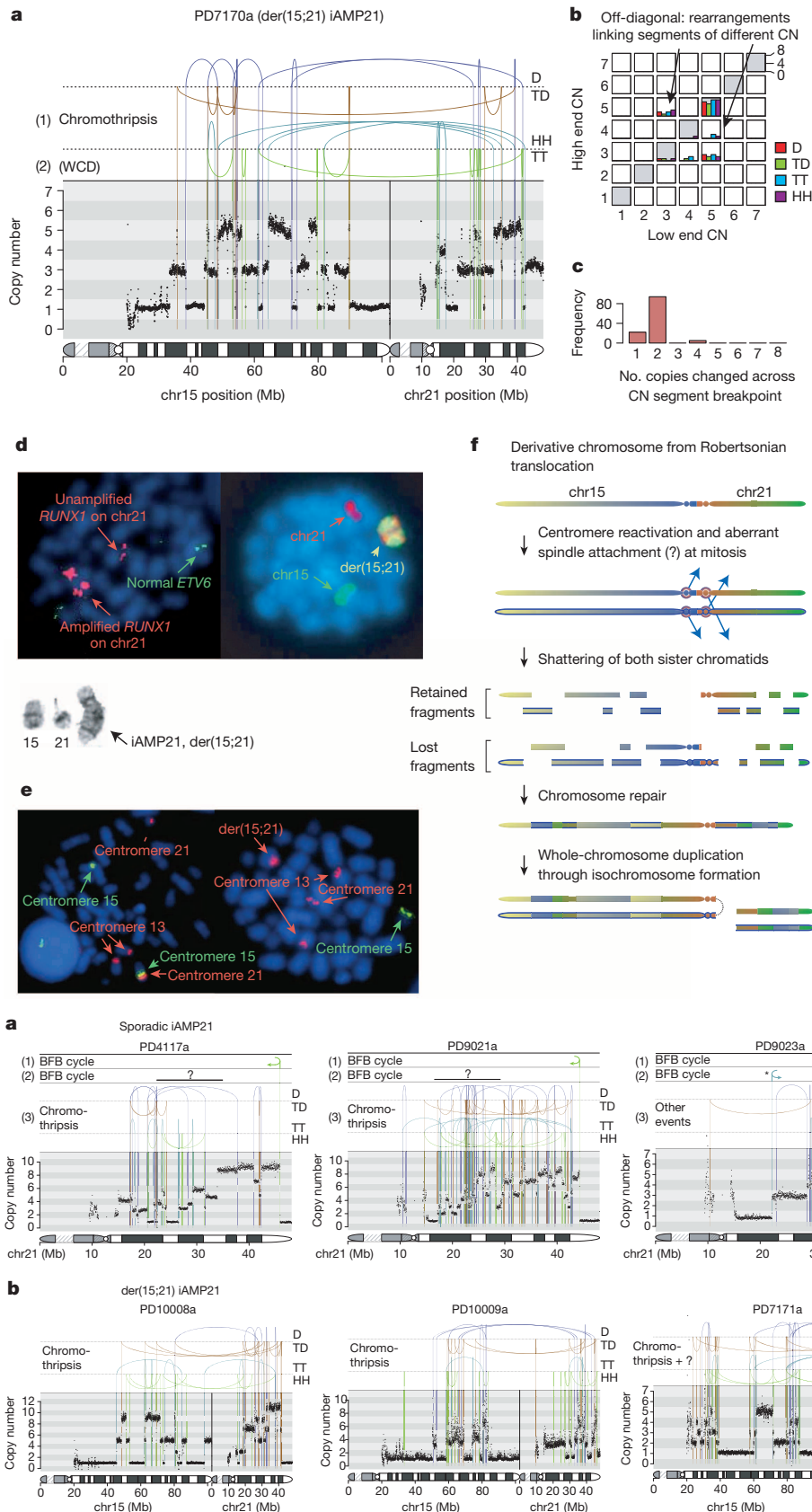


Figure 3 | Rearrangement patterns of the iAMP21 chromosome in the remaining patients. **a**, **b**, Rearrangement and copy-number patterns for chromosome 21 of sporadic iAMP21 ALL patients (**a**) and der(15;21) rearrangements in der(15;21) iAMP21 ALL patients (**b**). The inferred temporal orders of the major rearrangement events are shown with symbols (1), (2) and (3). In patients PD4117a and PD9021a, the fold-back rearrangement

Figure 2 | Rearrangements of der(15;21) in patient PD7170a. **a**, Rearrangement and copy-number pattern. The temporal order of the two major rearrangement events are marked (1) and (2). Rearrangements are separated based on their orientation: D, deletion-type; TD, tandem duplication-type; HH, head-to-head inverted; TT, tail-to-tail inverted. **b**, Copy-number jump distribution, showing the copy number at each end of each rearrangement. **c**, Copy-number step distribution, showing the distribution in magnitude of copy-number change at copy-number segmentation breakpoints. **d**, Representative metaphases. Left, several signals for *RUNX1* (large red signals) are clustered on two regions on the abnormal chromosome, and normal copies of *ETV6* on chromosome 12 (green). Right, cell painted for chromosomes 15 (green) and 21 (red). Inset shows partial G-banded karyotype: normal chromosome 15, normal chromosome 21 and isochromosome der(15;21). **e**, Left, cell shows representative metaphase from a non-leukaemic cell in patient PD10009a with rob(15;21)c, hybridized with centromere-specific probes for chromosomes 15 (green) and 13 and 21 (red), confirming that the Robertsonian chromosome is dicentric. Right, cell shows a leukaemia metaphase in which der(15;21) iAMP21 chromosome retains the chromosome 21 centromere (red), but not the chromosome 15 centromere (green). **f**, Model for evolution of iAMP21 in rob(15;21)c. Newly synthesized sister chromatids are indicated by a blue outline.

demarcating the second BFB repair breakpoint have probably been lost or obscured due to subsequent rearrangement events, and a '?' symbol is used to denote the uncertainty of their location. Inferred evolution of the derivative iAMP21 chromosomes are shown in the bottom panel. WC, whole chromosome; WCD, whole-chromosome duplication. Events with incomplete understanding are labelled '?'.

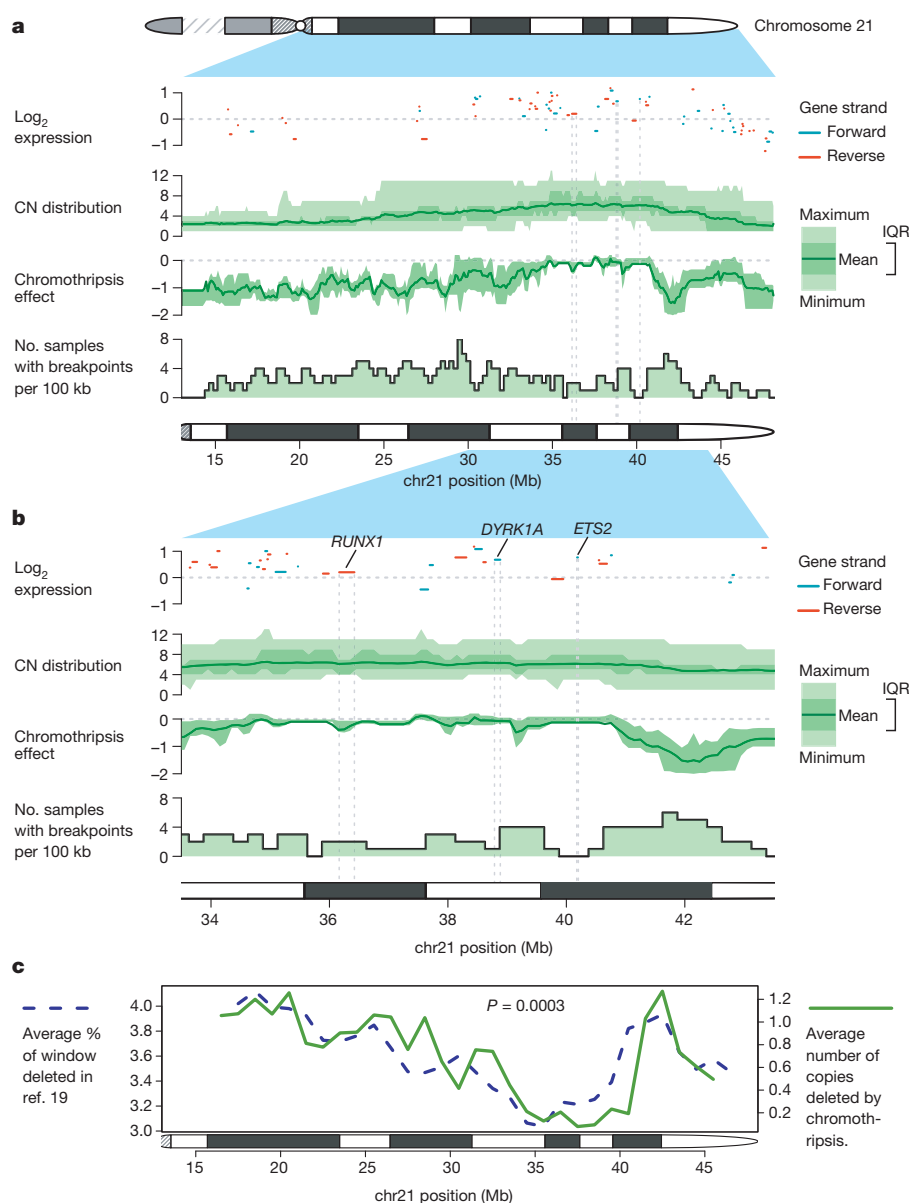


Figure 4 | Chromothripsis alters the copy-number landscape of chromosome 21 in a non-random fashion. **a**, **b**, Chromosome arm level (**a**) and zoomed-in view (**b**) of chromosome 21, showing gene expression, copy-number (CN) distribution, chromothripsis effect and distribution of

rearrangement breakpoints. In the gene expression panels, positive-strand genes are shown in blue and negative-strand genes are shown in red. **c**, Correlation between average rate of deletion in the ref. 19 data set and chromothripsis effect for chromosome 21. IQR, interquartile range.

der(15;21) iAMP21 cases, amplification was initiated by chromothripsis involving both sister chromatids of the rob(15;21)c chromosome. This seemed to be followed by further rearrangements in two cases and was completed by whole-chromosome duplications.

These data provide insight into why there is such specific enrichment of iAMP21 ALL in carriers of rob(15;21)c. Universally, the amplification is initiated by chromothripsis that affects both sister chromatids of the Robertsonian chromosome. This suggests that rob(15;21)c has a structural abnormality that specifically predisposes it, after replication, to the catastrophic shattering of chromothripsis. Using FISH, we demonstrated that the rob(15;21)c chromosome has centromeres from both chromosomes 15 and 21, and is thus dicentric (Fig. 2e and Extended Data Fig. 1). Our hypothesis, therefore, is that the two centromeres of the Robertsonian chromosome can occasionally confound attachment of mitotic spindles to the sister kinetochores, such that each chromatid connects to spindles emanating from opposite poles (Fig. 2f). During anaphase, this merotelic attachment would lead to lagging of both sister chromatids, rendering them jointly prone to chromothripsis. The der(15;21)

iAMP21 chromosomes consistently lose the chromosome 15 centromere, shown by FISH (Fig. 2e) and sequencing (Figs 2a and 3b), potentially enhancing stability of the derivative chromosome. In sporadic iAMP21 cases, chromothripsis frequently follows BFB cycles. Although we cannot know whether chromothripsis is an immediate consequence, it is plausible that the dicentric chromosome created by BFB repair could trigger chromothripsis, analogous to that seen with the dicentric rob(15;21)c^{17,18}.

The preceding analysis provides insight into the mutational processes shaping chromosome 21, but unless the resulting chromosome profile confers a selective advantage on the clone, it will not expand. We combined copy-number profiles and gene expression data³ from additional patients with iAMP21 ALL (Fig. 4). A consensus copy-number profile emerged in which regions from ~35.9–36.4 Mb and ~38.0–40.0 Mb of chromosome 21 were consistently the most highly amplified and over-expressed, including genes important in haematological malignancies such as *RUNX1*, *DYRK1A* and *ETS2*².

The final stage of iAMP21 generation usually involves duplication of the whole derivative chromosome, through whole-chromosome duplication,

isochromosome or ring formation. All duplications occurred after chromothripsis, suggesting that chromothripsis might be remodelling chromosome 21 in a non-random fashion. We used the inferred temporal evolution of somatic rearrangements to extract and average copy-number changes resulting from chromothripsis (Fig. 4). As expected, chromothripsis spared the most amplified regions, whereas on average one to two copies were deleted from other parts of the chromosome. Analysis of der(15;21) iAMP21 indicated that regions of chromosome 15 were also consistently lost or retained, although sample numbers are small (Extended Data Fig. 9).

Remarkably, the consensus chromothripsis landscape in iAMP21 closely mirrored the copy-number profile of chromosome 21 averaged over thousands of cancer samples across different cancer types^{19,20} ($P = 0.0003$; Fig. 4c; Supplementary Figs 27–30). This indicates that chromothripsis has a critical role in optimising the copy-number landscape of chromosome 21 to maximise the net selective advantage gained from subsequent rounds of whole-chromosome duplication.

From a detailed dissection of the mutational forces causing one particular subtype of one particular cancer, findings with general significance have emerged. Carriers of constitutional rob(15;21)c chromosomes are specifically but highly predisposed to iAMP21 ALL. Usually, constitutional risk of cancer is mediated by variation in coding sequence or gene regulation, but here it seems to be transmitted through a propensity for the Robertsonian chromosome to undergo chromothripsis after replication. This may be because it is dicentric and prone to anaphase bridging, which would dovetail with the frequent occurrence of chromothripsis following BFB cycles in sporadic iAMP21 ALL. This hypothesis is consistent with the finding that lagging chromosomes during anaphase can become sequestered in micronuclei and subjected to chromosomal pulverisation before rejoining the main nucleus^{17,18}. More generally, the study of iAMP21 ALL has illustrated how large-scale copy-number changes can be optimised by spatially and temporally coordinated genomic instability taking several complementary forms. BFB cycles can generate rapid-fire, focal amplification; chromothripsis causes loss of multiple, non-contiguous chromosomal regions; and whole-chromosome duplication gives expansive, low-amplitude amplification. Their combined activity, therefore, gives considerably more flexibility to shaping large-scale chromosomal copy-number profiles than any one process alone. Of course, clones sample these mutational processes randomly, so only when the aggregate fitness of such changes is positive will the clone have the selective advantage to expand.

METHODS SUMMARY

Information was available from 21 iAMP21 patients and 12 with iAMP21 and rob(15;21)(q10;q10)c (Supplementary Table 2). Paired-end sequencing data were generated as 37–75 bp paired reads from 400–500 bp fragments as previously described¹⁹. The deductive reasoning for reconstructing temporal evolution of complex rearrangements followed principles formulated previously²¹. Confirmatory PCR across the breakpoints was performed for the vast majority of identified rearrangements (Supplementary Results).

Received 13 September 2013; accepted 30 January 2014.

Published online 23 March 2014.

1. Stratton, M. R., Campbell, P. J. & Futreal, P. A. The cancer genome. *Nature* **458**, 719–724 (2009).
2. Rand, V. et al. Genomic characterization implicates iAMP21 as a likely primary genetic event in childhood B-cell precursor acute lymphoblastic leukemia. *Blood* **117**, 6848–6855 (2011).
3. Strefford, J. C. et al. Complex genomic alterations and gene expression in acute lymphoblastic leukemia with intrachromosomal amplification of chromosome 21. *Proc. Natl Acad. Sci. USA* **103**, 8167–8172 (2006).
4. Stiller, C. A., Kroll, M. E., Boyle, P. J. & Feng, Z. Population mixing, socioeconomic status and incidence of childhood acute lymphoblastic leukaemia in England and Wales: analysis by census ward. *Br. J. Cancer* **98**, 1006–1011 (2008).
5. Moorman, A. V. et al. Prognosis of children with acute lymphoblastic leukemia (ALL) and intrachromosomal amplification of chromosome 21 (iAMP21). *Blood* **109**, 2327–2330 (2007).

6. Moorman, A. V. et al. Risk-directed treatment intensification significantly reduces the risk of relapse among children and adolescents with acute lymphoblastic leukemia and intrachromosomal amplification of chromosome 21: a comparison of the MRC ALL97/99 and UKALL2003 trials. *J. Clin. Oncol.* **31**, 3389–3396 (2013).
7. Heerema, N. A. et al. Intrachromosomal amplification of chromosome 21 is associated with inferior outcomes in children with acute lymphoblastic leukemia treated in contemporary standard-risk children's oncology group studies: a report from the children's oncology group. *J. Clin. Oncol.* **31**, 3397–3402 (2013).
8. Hamerton, J. L., Canning, N., Ray, M. & Smith, S. A cytogenetic survey of 14,069 newborn infants. I. Incidence of chromosome abnormalities. *Clin. Genet.* **8**, 223–243 (1975).
9. Jacobs, P. A., Browne, C., Gregson, N., Joyce, C. & White, H. Estimates of the frequency of chromosome abnormalities detectable in unselected newborns using moderate levels of banding. *J. Med. Genet.* **29**, 103–108 (1992).
10. Campbell, P. J. et al. Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nature Genet.* **40**, 722–729 (2008).
11. Campbell, P. J. et al. The patterns and dynamics of genomic instability in metastatic pancreatic cancer. *Nature* **467**, 1109–1113 (2010).
12. Sinclair, P. B. et al. Analysis of a breakpoint cluster reveals insight into the mechanism of intrachromosomal amplification in a lymphoid malignancy. *Hum. Mol. Genet.* **20**, 2591–2602 (2011).
13. Robinson, H. M., Harrison, C. J., Moorman, A. V., Chudoba, I. & Strefford, J. C. Intrachromosomal amplification of chromosome 21 (iAMP21) may arise from a breakage-fusion-bridge cycle. *Genes Chromosomes. Cancer* **46**, 318–326 (2007).
14. Korb, J. O. & Campbell, P. J. Criteria for inference of chromothripsis in cancer genomes. *Cell* **152**, 1226–1236 (2013).
15. Rausch, T. et al. Genome sequencing of pediatric medulloblastoma links catastrophic DNA rearrangements with TP53 mutations. *Cell* **148**, 59–71 (2012).
16. Stephens, P. J. et al. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* **144**, 27–40 (2011).
17. Crasta, K. et al. DNA breaks and chromosome pulverization from errors in mitosis. *Nature* **482**, 53–58 (2012).
18. Hatch, E. M., Fischer, A. H., Deerinc, T. J. & Hetzer, M. W. Catastrophic nuclear envelope collapse in cancer cell micronuclei. *Cell* **154**, 47–60 (2013).
19. Beroukhi, R. et al. The landscape of somatic copy-number alteration across human cancers. *Nature* **463**, 899–905 (2010).
20. Kim, T. M. et al. Functional genomic analysis of chromosomal aberrations in a compendium of 8000 cancer genomes. *Genome Res.* **23**, 217–227 (2013).
21. Greenman, C. D. et al. Estimation of rearrangement phylogeny for cancer genomes. *Genome Res.* **22**, 346–361 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank member laboratories of the United Kingdom Cancer Cytogenetic Group (UKCCG) for providing cytogenetic data and material. Primary childhood leukaemia samples used in this study were provided by the Leukaemia and Lymphoma Research Childhood Leukaemia Cell Bank working with the laboratory teams in the Bristol Genetics Laboratory, Southmead Hospital, Bristol, UK; Molecular Biology Laboratory, Royal Hospital for Sick Children, Glasgow, UK; Molecular Haematology Laboratory, Royal London Hospital, London, UK; Molecular Genetics Service and Sheffield Children's Hospital, Sheffield, UK. We also thank all the members of the NCRI Childhood Cancer and Leukaemia Group (CCLG) Leukaemia Subgroup for access to material and data on clinical trial patients. This work was supported by the Wellcome Trust (077012/Z/05/Z); Leukaemia and Lymphoma Research Specialist Programme and European Research Council (249891). P.J.C. has a Wellcome Trust Senior Clinical Research Fellowship (WT088340MA). P.V.L. is supported by a postdoctoral research fellowship and P.V. is a Senior Clinical Investigator funded by the Research Foundation – Flanders (F.W.O.). P.S. is funded by the European Research Council (grant 249891).

Author Contributions C.J.H. and P.J.C. designed the study; Y.L. carried out and interpreted the sequencing and associated analysis, assisted by E.P. and P.J.S.; C.S. and S.R. coordinated the study; C.S. carried out the FISH analyses and interpreted the FISH and SNP6.0 results; S.R. carried out the initial sequence analysis and associated validation; B.D.Y. assisted with the analysis of SNP6.0 data; C.S. and H.M.R. interpreted the cytogenetic findings; O.J., B.R., and M.M. performed laboratory analyses; P.J., M.G., P.T., N.B., N.T., C.H. and L.G. provided data on incidence of rob(15;21)c cases; P.J., F.M.R., N.A.H., A.J.C., N.B., N.T., M.R.T., S.D., J.B., N.D. and P.V. provided rob(15;21)c cases and associated clinical and genetic data to be included in the study; A.V.M. and R.J.Q.M. provided the incidence data and calculated the relative risk values; P.S. and V.R. provided data interpretation; J.C. and P.V.L. ran copy-number analyses and coordinated analysis of publicly available solid tumour cancer data; M.R.S. contributed to the analysis and interpretation of the sequencing studies. P.J.C. and C.J.H. assimilated the data and wrote the manuscript, with support from all authors.

Author Information Genome sequence data have been deposited at the European Genome-Phenome Archive (<http://www.ebi.ac.uk/ega/>), hosted by the EBI with accession number EGAD00001000658 (<https://www.ebi.ac.uk/ega/datasets/EGAD00001000658>). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.J.H. (christine.harrison@newcastle.ac.uk) or P.J.C. (pc8@sanger.ac.uk).

XBP1 promotes triple-negative breast cancer by controlling the HIF1 α pathway

Xi Chen^{1,2}, Dimitrios Iliopoulos^{3,4*}, Qing Zhang^{5*}, Qianzi Tang^{6,7*}, Matthew B. Greenblatt⁸, Maria Hatziaepostolou^{3,4}, Elgene Lim⁹, Wai Leong Tam¹⁰, Min Ni⁹, Yiwen Chen¹¹, Junhua Mai¹², Haifa Shen^{12,13}, Dorothy Z. Hu¹⁴, Stanley Adoro^{1,2}, Bella Hu¹⁵, Minkyung Song^{1,2}, Chen Tan^{1,2}, Melissa D. Landis¹⁶, Mauro Ferrari^{2,12}, Sandra J. Shin¹⁷, Myles Brown⁹, Jenny C. Chang^{2,16}, X. Shirley Liu¹¹ & Laurie H. Glimcher^{1,2}

Cancer cells induce a set of adaptive response pathways to survive in the face of stressors due to inadequate vascularization¹. One such adaptive pathway is the unfolded protein (UPR) or endoplasmic reticulum (ER) stress response mediated in part by the ER-localized transmembrane sensor IRE1 (ref. 2) and its substrate XBP1 (ref. 3). Previous studies report UPR activation in various human tumours^{4–6}, but the role of XBP1 in cancer progression in mammary epithelial cells is largely unknown. Triple-negative breast cancer (TNBC)—a form of breast cancer in which tumour cells do not express the genes for oestrogen receptor, progesterone receptor and HER2 (also called ERBB2 or NEU)—is a highly aggressive malignancy with limited treatment options^{7,8}. Here we report that XBP1 is activated in TNBC and has a pivotal role in the tumorigenicity and progression of this human breast cancer subtype. In breast cancer cell line models, depletion of *XBP1* inhibited tumour growth and tumour relapse and reduced the CD44^{high}CD24^{low} population. Hypoxia-inducing factor 1 α (HIF1 α) is known to be hyperactivated in TNBCs^{9,10}. Genome-wide mapping of the XBP1 transcriptional regulatory network revealed that XBP1 drives TNBC tumorigenicity by assembling a transcriptional complex with HIF1 α that regulates the expression of HIF1 α targets via the recruitment of RNA polymerase II. Analysis of independent cohorts of patients with TNBC revealed a specific *XBP1* gene expression signature that was highly correlated with HIF1 α and hypoxia-driven signatures and that strongly associated with poor prognosis. Our findings reveal a key function for the XBP1 branch of the UPR in TNBC and indicate that targeting this pathway may offer alternative treatment strategies for this aggressive subtype of breast cancer.

We determined UPR activation status in several breast cancer cell lines. XBP1 expression was readily detected in both luminal and basal-like breast cancer cell lines, but the level of its spliced form was higher in the latter which consist primarily of TNBC cells; XBP1 activation was also higher in primary TNBC patient samples (Fig. 1a, b). PERK but not ATF6 was also activated (Extended Data Fig. 1a), and transmission electron microscopy revealed more abundant and dilated ER in multiple TNBC cell lines (Extended Data Fig. 1b). These data reveal a state of basal ER stress in TNBC cells.

XBP1 silencing impaired soft agar colony-forming ability and invasiveness (Extended Data Fig. 1c) of multiple TNBC cell lines, indicating that XBP1 regulates TNBC anchorage-independent growth and invasiveness. We next used an orthotopic xenograft mouse model with inducible

expression of two *XBP1* short hairpin RNAs (shRNAs) in MDA-MB-231 cells. Tumour growth and metastasis to lung were significantly inhibited by *XBP1* shRNAs (Fig. 1c–e and Extended Data Fig. 1d–g). This was not due to altered apoptosis (caspase 3), cell proliferation (Ki67) or hyperactivation of IRE1 and other UPR branches (Fig. 1e and Extended Data Fig. 1h, i). Instead, XBP1 depletion impaired angiogenesis as demonstrated by the presence of fewer intratumoral blood vessels (CD31 staining) (Fig. 1e). Subcutaneous xenograft experiments using two other TNBC cell lines confirmed our findings (Extended Data Fig. 1j, k). Notably, XBP1 silencing in a patient-derived TNBC xenograft model (BCM-2147) significantly decreased tumour incidence (Fig. 1f and Extended Data Fig. 1l, m).

TNBC patients have the highest rate of relapse within 1–3 years despite adjuvant chemotherapy^{7,8}. To examine XBP1's effect on tumour relapse after chemotherapeutic treatment, we treated MDA-MB-231 xenograft-bearing mice with doxorubicin and *XBP1* shRNA. Notably, combination treatment not only blocked tumour growth but also inhibited or delayed tumour relapse (Fig. 2a).

Tumour cells expressing CD44^{high}CD24^{low} have been shown to mediate tumour relapse in some instances^{11–13}. To test whether XBP1 targeted the CD44^{high}CD24^{low} population, we examined the mammosphere-forming ability of cells derived from treated tumours (day 20). Mammosphere formation was increased in doxorubicin-treated tumour cells, whereas tumours treated with doxorubicin plus *XBP1* shRNA displayed substantially reduced mammosphere formation (Fig. 2b), a finding confirmed using another chemotherapeutic agent, paclitaxel (Extended Data Fig. 2a, b). Hypoxia activates the UPR, and *XBP1* knockdown also markedly reduced mammosphere formation in hypoxic conditions (Extended Data Fig. 2b). Furthermore, CD44 expression was reduced in XBP1-depleted tumours (Extended Data Fig. 2c).

To interrogate XBP1's effect on CD44^{high}CD24^{low} cell function further, we used mammary epithelial cells (MCF10A) carrying an inducible Src oncogene (ER-Src), where v-Src is fused with the oestrogen receptor ligand-binding domain¹⁴. Tamoxifen treatment results in neoplastic transformation and gain of a CD44^{high}CD24^{low} population that has been previously associated with tumour-initiating properties¹⁵. In transformed MCF10A-ER-Src cells, XBP1 splicing was increased in the CD44^{high}CD24^{low} population (Fig. 2c), whereas *XBP1* silencing reduced the CD44^{high}CD24^{low} fraction (Extended Data Fig. 2d, e) and markedly suppressed mammosphere formation (Extended Data Fig. 2f), phenotypes

¹Sandra and Edward Meyer Cancer Center of Weill Cornell Medical College, 1300 York Avenue, New York, New York 10065, USA. ²Department of Medicine, Weill Cornell Medical College, 1300 York Avenue, New York, New York 10065, USA. ³Center for Systems Biomedicine, Division of Digestive Diseases, David Geffen School of Medicine, University of California, Los Angeles, California 90095, USA. ⁴Department of Cancer Immunology and AIDS, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA. ⁵Lineberger Comprehensive Cancer Center, Department of Pathology and Laboratory Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. ⁶Department of Bioinformatics, School of Life Science and Technology, Tongji University, Shanghai 200092, China. ⁷Institute of Animal Genetics and Breeding, College of Animal Science and Technology, Sichuan Agricultural University, Ya'an, Sichuan 625014, China. ⁸Department of Pathology, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA. ⁹Department of Medical Oncology, Dana-Farber Cancer Institute and Department of Medicine, Harvard Medical School, Boston, Massachusetts 02115, USA. ¹⁰Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA. ¹¹Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute and Harvard School of Public Health, Boston, Massachusetts 02215, USA. ¹²Department of Nanomedicine, Houston Methodist Research Institute, Houston, Texas 77030, USA. ¹³Department of Cell and Developmental Biology, Weill Cornell Medical College, 1300 York Avenue, New York, New York 10065, USA. ¹⁴Endocrine Unit, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. ¹⁵Division of Hematology/Oncology, Children's Hospital Boston, Boston, Massachusetts 02115, USA. ¹⁶Houston Methodist Cancer Center, Houston, Texas 77030, USA. ¹⁷Department of Pathology and Laboratory Medicine, Weill Cornell Medical College, 1300 York Avenue, New York, New York 10065, USA.

*These authors contributed equally to this work.

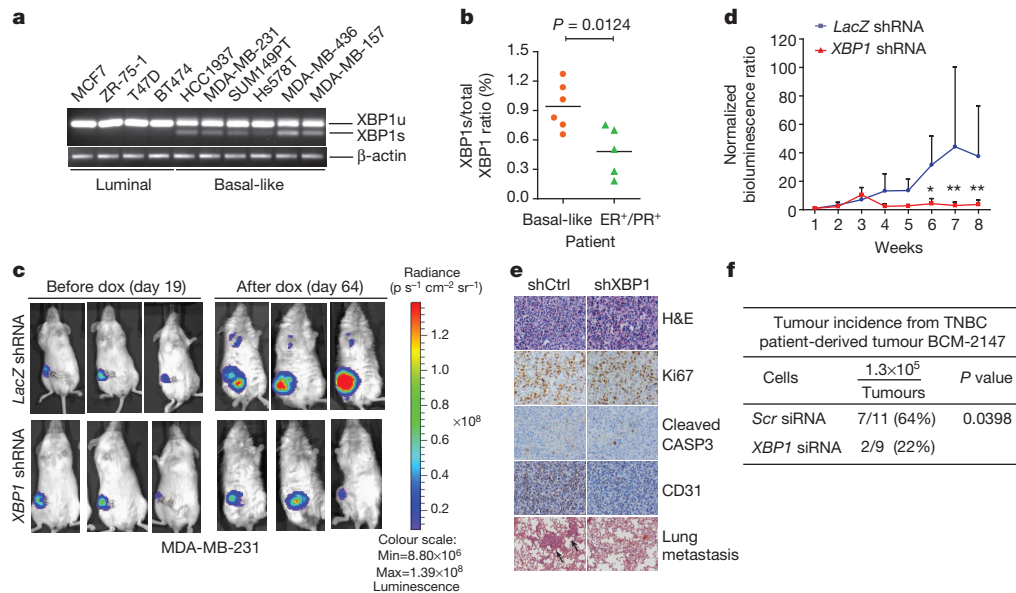


Figure 1 | *XBP1* silencing blocks TNBC cell growth and invasiveness.

a, b, RT-PCR analysis of XBP1 splicing in luminal and basal-like cell lines (**a**) or primary tissues from 6 TNBC patients and 5 oestrogen/progesterone-positive (ER/PR⁺) patients (**b**). XBP1u, unspliced XBP1; XBP1s, spliced XBP1. β -actin was used as loading control. **c**, Representative bioluminescent images of orthotopic tumours formed by MDA-MB-231 cells as in Extended Data Fig. 1d. Bioluminescent images were obtained 5 days after transplantation and serially after mice were begun on chow containing doxycycline (day 19) for 8 weeks.

not attributable to a direct effect of XBP1 on cell viability (Extended Data Fig. 2g, h). Furthermore, limiting dilution experiments demonstrated loss of tumour-seeding ability in *XBP1*-depleted cells (Fig. 2d). CD44^{high} CD24^{low} cells sorted from TNBC patient samples confirmed increased XBP1 splicing and other UPR markers, and *XBP1* silencing impaired mammosphere-forming ability (Fig. 2e, f and Extended Data Fig. 3a). Conversely, overexpression of spliced XBP1 (XBP1s) in CD44^{low} CD24^{high} cells resulted in gain of mammosphere-forming ability and increased resistance to doxorubicin treatment (Extended Data Fig. 3b, c). Notably, patient-derived CD44^{low} CD24^{high} cells overexpressing XBP1s, but not control parental cells, initiated tumour formation in immunodeficient

mice (Extended Data Fig. 3d, e). These data establish a critical role of XBP1 in CD44^{high} CD24^{low} cells within TNBC.

Chromatin immunoprecipitation coupled with ultra-high-throughput DNA sequencing (ChIP-seq) and motif analysis of XBP1 in MDA-MB-231 cells revealed statistically significant enrichment of both the HIF1 α and XBP1 motifs (Fig. 3a and Extended Data Fig. 4a), indicating frequent co-localization of HIF1 α and XBP1 to the same regulatory elements. HIF1 α is hyperactivated in TNBCs, required for the maintenance of CD44^{high} CD24^{low} cells^{9,10,16,17}, and is regulated in response to micro-environmental oxygen levels. XBP1 ChIP-seq was therefore also carried out in MDA-MB-231 and Hs578T cells cultured under hypoxia and glucose deprivation conditions for 24 h. Exposure to these stressors increased XBP1 splicing, resulting in a corresponding increase in signal intensity (Extended Data Fig. 4b–f) and further enrichment of HIF1 α motifs in TNBC (Fig. 3a), but interestingly not in luminal breast cancer cells (Extended Data Fig. 4g).

HIF1 α motif enrichment in the XBP1 ChIP-seq data set indicated that XBP1 and HIF1 α might interact within the same transcriptional complex.

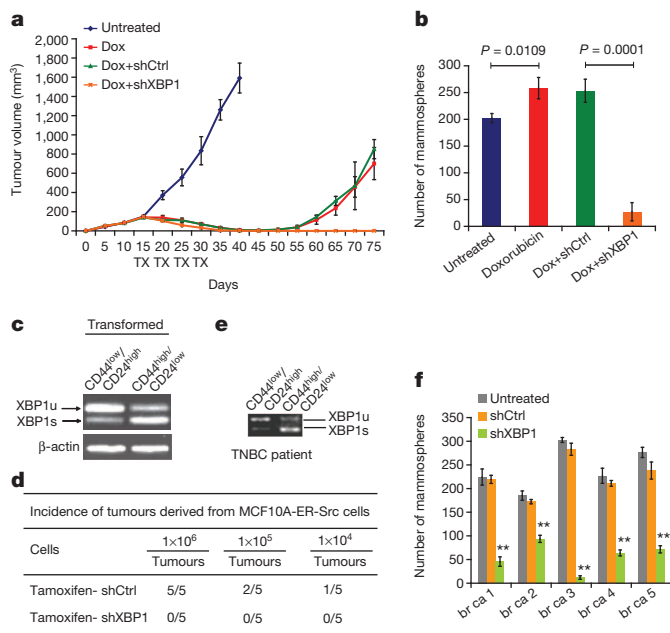


Figure 2 | *XBP1* is required for tumour relapse and CD44^{high}CD24^{low} cells.

a, Tumour growth of MDA-MB-231 cells untreated or treated with doxorubicin (dox), or doxorubicin plus control shRNA (shCtrl), or doxorubicin plus *XBP1* shRNA (shXBP1) in athymic nude mice. Data are shown as mean \pm s.d. of biological replicates ($n = 5$). TX indicates treatment. **b**, Number of mammospheres per 1,000 cells generated from day 20 xenograft tumours under different treatments as indicated. Data are shown as mean \pm s.d. of biological replicates ($n = 3$). **c**, RT-PCR analysis of XBP1 splicing in tamoxifen-treated CD44^{low} CD24^{high} and CD44^{high} CD24^{low} cells. **d**, The indicated number of tamoxifen-treated MCF10A-ER-Src cells infected with control shRNA or *XBP1* shRNA were injected into NOD/SCID/*Il2r*^{-/-} mice and the tumour incidence reported at 12 weeks after transplantation. **e**, RT-PCR analysis of XBP1 splicing in CD44^{low} CD24^{high} and CD44^{high} CD24^{low} cells purified from a TNBC patient. **f**, Number of mammospheres per 1,000 cells generated from primary tissue samples from five patients with TNBC (br ca 1–5) that were untreated or infected with lentiviruses encoding control shRNA or *XBP1* shRNA. Data are shown as mean \pm s.d. of technical replicates.

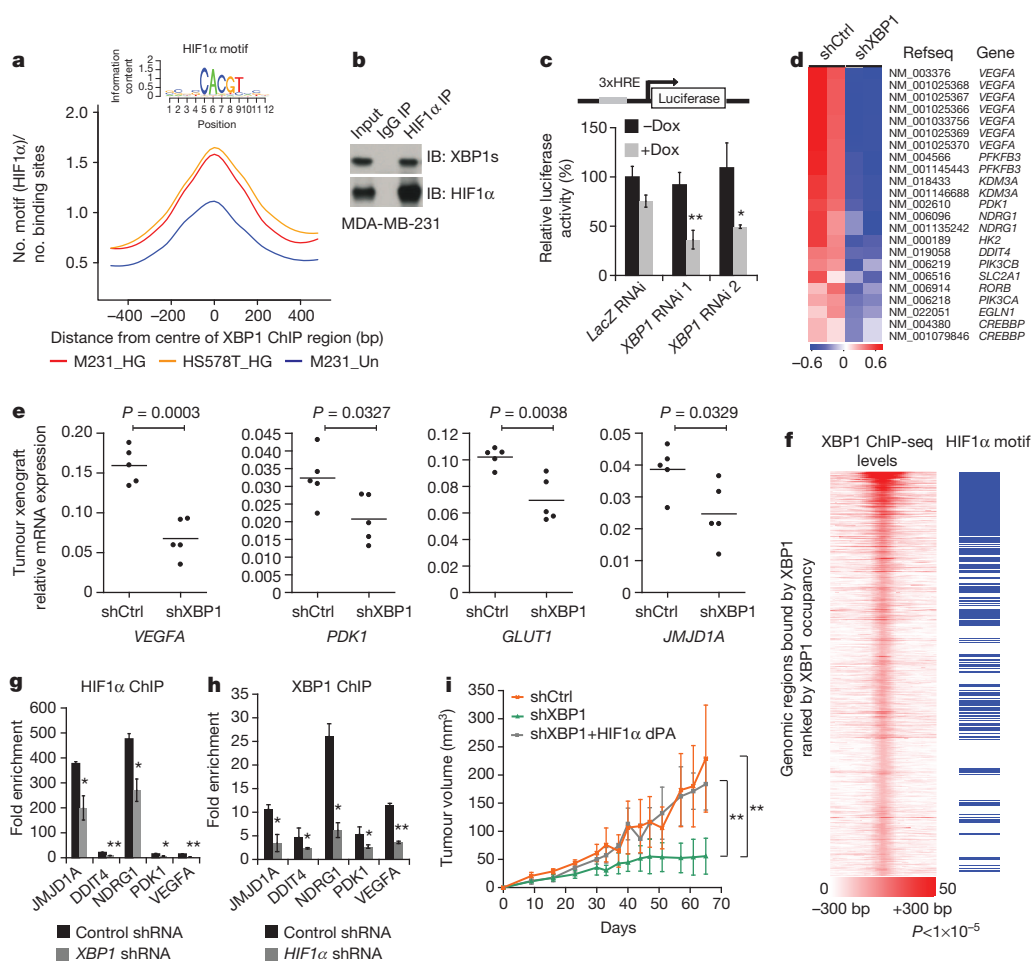


Figure 3 | HIF1 α is a co-regulator of XBP1. **a**, Motif enrichment analysis in the XBP1 binding sites in untreated (Un) or stressed (0.1% O₂ and glucose deprivation; HG) MDA-MB-231 or Hs578T cells. The 1-kilobase (kb) region surrounding the summit of the XBP1 peak is equally divided into 50 bins. The average HIF1 α motif occurrence over top 1,000 XBP1 peaks in each bin is plotted. The sequence logo for HIF1 α motif is shown as an inset. The corresponding *P* values of each condition are: M231_Un, 7.78×10^{-20} ; M231_HG, $<1.08 \times 10^{-30}$; Hs578T_HG, $<1.08 \times 10^{-30}$. **b**, Nuclear extracts from MDA-MB-231 cells (treated with 0.1% O₂ for 10 h first and then with 1 μ g ml⁻¹ tunicamycin for another 6 h in 0.1% O₂) were subjected to co-immunoprecipitation with anti-HIF1 α antibody or rabbit IgG. IB, immunoblot. **c**, Schematic of the luciferase reporter constructs. 3 \times HRE reporter was co-transfected with doxycycline (dox)-inducible constructs encoding control or two XBP1 shRNAs into MDA-MB-231 cells. Cells were treated in 0.1% O₂ for 24 h and luciferase activity assayed. Experiments were performed in triplicate and data are shown as mean \pm s.d. **d**, Gene expression microarray heat map showing that genes involved in the HIF1 α -mediated hypoxia response were differentially expressed after XBP1 knockdown.

e, RT-PCR analysis of HIF1 α target gene expression after knockdown of XBP1 in MDA-MB-231-derived xenograft tumours in NOD/SCID/IL2r $\gamma^{-/-}$ mice. Results are presented relative to β -actin expression. *n* = 5. **f**, Plot showing the genome-wide association between the strength of XBP1 binding and the occurrence of the HIF1 α motif. The signal of XBP1 ChIP-seq peaks is shown as a heat map using red (the strongest signal) and white (the weakest signal) colour scheme. Each row shows \pm 300 bp centred on the XBP1 ChIP-seq peak summits. Rows are ranked by XBP1 occupancy. The horizontal blue lines denote the presence of the HIF1 α motif. **g**, **h**, Chromatin extracts from control, XBP1 knockdown (**g**) or HIF1 α knockdown (**h**) MDA-MB-231 cells (treated with 0.1% O₂ for 24 h) were subjected to ChIP using anti-HIF1 α (**g**) or anti-XBP1 antibodies (**h**). Data are shown as mean \pm s.d. of technical triplicates. Results show a representative of two independent experiments. **i**, Growth curve of tumours formed by MDA-MB-231 cells infected with inducible control shRNA (shCtrl), XBP1 shRNA (shXBP1) or XBP1 shRNA plus constitutively activated HIF1 α (shXBP1 + HIF1 α dPA). Mice were fed with doxycycline chow from day 7. Data are shown as mean \pm s.d. of biological replicates (*n* = 5). **P* < 0.05, ***P* < 0.01.

Co-immunoprecipitation experiments revealed a physical interaction of HIF1 α , but not HIF2 α , with XBP1 in 293T cells co-expressing HIF1 and XBP1s cultured under hypoxic conditions; an interaction was also observed with endogenous proteins in two TNBC cell lines: MDA-MB-231 and Hs578T (Fig. 3b and Extended Data Fig. 4h–j). Subcellular fractionation revealed that this interaction occurs in the nucleus, and that unspliced XBP1u protein was not detectable (Extended Data Fig. 4k, l). Glutathione *S*-transferase (GST) pull-down experiments showed that HIF1 α interacts with the XBP1s amino-terminus b-zip domain (Extended Data Fig. 4m, n).

We next established that XBP1 and HIF1 α co-occupied several well-known HIF1 α targets using ChIP-qPCR (Extended Data Fig. 5a–c). ChIP-re-ChIP assays using anti-XBP1s followed by anti-HIF1 α antibodies

confirmed that XBP1s and HIF1 α simultaneously co-occupy these common targets (Extended Data Fig. 5d). DNA pull-down assays with a HIF1 α (ref. 18) specific probe precipitated XBP1s in MDA-MB-231 nuclear extracts under hypoxia, indicating their presence in the same complex (Extended Data Fig. 5e, f). XBP1 depletion by two independent shRNA constructs markedly reduced hypoxia response element (HRE) luciferase activity under hypoxia (Fig. 3c). Conversely, XBP1s expression dose-dependently transactivated the HRE reporter (Extended Data Fig. 5g, h), confirming that XBP1 augments HIF1 α activity.

When we profiled the differential transcriptome induced by XBP1 silencing in MDA-MB-231 cells, gene-set enrichment analysis identified significant enrichment of HIF1 α -mediated hypoxia response pathway genes (Fig. 3d and Extended Data Fig. 6a). XBP1 depletion downregulated

HIF1 α targets *VEGFA*, *PDK1*, *GLUT1* (also called *SLC2A1*) and *DDIT4* expression in both normoxic and hypoxic conditions (Extended Data Fig. 6b), and these results were validated in breast cancer xenografts (Fig. 3e) and Hs578T cells (Extended Data Fig. 6c). However, XBP1 depletion in luminal tumours did not affect these targets (Extended Data Fig. 6d).

To explore the consequences of this cooperation further, we examined how XBP1 or HIF1 α loss affected the transcription of common target genes. We found that high XBP1 occupancy at its binding sites was associated with increased occurrence of the HIF1 α motif across the genome in TNBC (Fig. 3f). Whereas XBP1 depletion had no immediate effect on HIF1 α expression, it substantially attenuated concurrent HIF1 α and RNA polymerase II occupancy (Extended Data Fig. 6e–g

and Fig. 3g). Similarly, XBP1 and RNA polymerase II occupancy at co-bound sites was likewise reduced in the absence of HIF1 α under hypoxic conditions (Fig. 3h and Extended Data Fig. 6h–i). These results indicate that the assembly of the XBP1–HIF1 α complex on target promoters is crucial for their transcription, via the recruitment of RNA polymerase II.

To establish whether HIF1 α contributes to the function of XBP1 in TNBC, we performed rescue experiments using a haemagglutinin (HA)-tagged constitutively activated hydroxylation-mutant HIF1 α construct (HA–HIF1 α dPA: P402A/P564A). XBP1 splicing was not directly regulated by HIF1 α (Extended Data Fig. 7a–c). Enforced overexpression of HIF1 α dPA in XBP1-depleted cells restored expression of HIF1 α targets and rescued anchorage-independent growth, mammosphere-forming ability, angiogenesis and *in vivo* tumour growth (Fig. 3i and Extended Data Fig. 7d–h). Conversely, HIF1 α silencing in XBP1s overexpressing cells markedly compromised their ability to sustain mammosphere formation (Extended Data Fig. 7i). Hypoxia is a physiological UPR inducer in cancer¹⁹ and XBP1s co-localizes with hypoxia marker CA9 in tumours²⁰. Our experiments demonstrate that XBP1 functions to sustain the hypoxia response via regulating the HIF1 α transcriptional program (Extended Data Fig. 8), which ensures maximum HIF activity and adaptive responses to the cytotoxic microenvironment of solid tumours.

Integrated analysis of XBP1 ChIP-seq data and gene expression profiles identified 96 genes directly bound and upregulated by XBP1. This gene set was defined as the XBP1 signature (Fig. 4a and Supplementary Table 1). Its expression was highly correlated with hypoxia-driven signatures in TNBC (Pearson's correlation coefficient = 0.61; $P = 2.28 \times 10^{-60}$), but not in ER⁺ breast cancer patients (coefficient = 0.03; $P = 0.64$) (Extended Data Fig. 9a, b). Survival analysis using an aggregate breast cancer data set for 193 TNBC patient samples²¹ demonstrated that tumours with an elevated XBP1 signature displayed shorter relapse-free survival (log-rank test, $P = 0.00677$) (Fig. 4b). Cox regression analysis showed that association of the signature with relapse-free survival remained significant after controlling for tumour size, grade and chemotherapy treatment ($P = 0.00453$, Supplementary Table 2). These findings were validated in a separate cohort of 190 TNBC patients (Fig. 4c). Importantly, the XBP1 signature did not correlate with clinical outcome of ER⁺ breast cancer patients ($P = 0.553$) (Extended Data Fig. 9c), indicating its specific prognostic value for TNBC. Expression of the XBP1-regulated HIF1 α program was also associated with decreased relapse-free survival only in TNBC ($P = 0.00911$) (Fig. 4d and Extended Data Fig. 9d). Although XBP1 silencing also affects luminal breast cancer growth, it does so via a mechanism not involving HIF1 α (Extended Data Figs 4g, 6d and 10).

In conclusion, we have uncovered a key function of XBP1 in the tumorigenicity, progression and recurrence of TNBC, and have identified XBP1's control of the HIF1 α transcriptional program as the major mechanism. XBP1 pathway activation correlates with poor patient survival in TNBC patients, indicating that UPR inhibitors in combination with standard chemotherapy may improve the effectiveness of anti-tumour therapies.

METHODS SUMMARY

Orthotopic tumour growth assays. Female NOD/SCID/IL2r $\gamma^{-/-}$ or nude mice (Taconic) were injected orthotopically with 1.5×10^6 viable tumour cells re-suspended in 40 μ l Matrigel (BD Biosciences) into mammary glands and fed chow containing 6 g doxycycline per kg (Bioserv). For bioluminescent detection, mice were given a single intraperitoneal injection of luciferin, ketamine and xylazine and imaged with an IVIS imaging camera (Xenogen). Imaging intensity was normalized to the luminescence signal of individual mice before doxycycline chow treatment. The average luminescence ratio of treatment groups (*lacZ* or *XBP1* shRNA) was plotted over the course of doxycycline chow treatment and results presented as mean \pm s.d.

Mammosphere formation assay. Mammospheres were generated from cells in suspension ($1,000$ cells ml^{-1}) in serum-free DMEM/F12 media, supplemented with B27 (1:50, Invitrogen), 0.4% BSA, 20 ng ml^{-1} EGF and 4 μg ml^{-1} insulin. After 6 days mammospheres were typically >75 μm in size with $\sim 97\%$ CD44^{high}CD24^{low}. For serial passaging, 6-day-old mammospheres were collected, dissociated to single cells with trypsin and re-grown in suspension for 6 days.

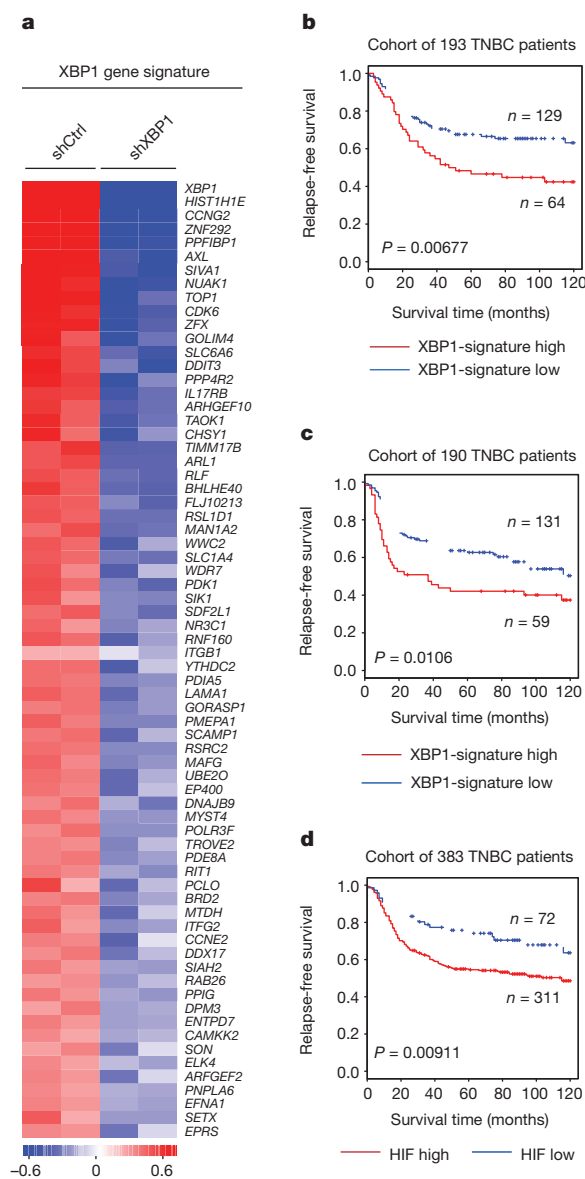


Figure 4 | XBP1 genetic signature is associated with human TNBC prognosis. **a**, Heat map showing the expression profile of genes bound by XBP1 and differentially expressed after XBP1 knockdown. **b**, **c**, Kaplan–Meier graphs demonstrating a significant association between elevated expression of the XBP1 signature (red line) and shorter relapse-free survival in two cohorts of patients with TNBC (**b** and **c**). **d**, Kaplan–Meier graphs showing significant association of elevated HIF1 α gene signature expression (red line) with shorter relapse-free survival in a cohort of 383 TNBC patients. The log-rank test P values are shown.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 21 November 2012; accepted 31 January 2014.

Published online 23 March 2014.

1. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).
2. Walter, P. & Ron, D. The unfolded protein response: from stress pathway to homeostatic regulation. *Science* **334**, 1081–1086 (2011).
3. Yoshida, H., Matsui, T., Yamamoto, A., Okada, T. & Mori, K. XBP1 mRNA is induced by ATF6 and spliced by IRE1 in response to ER stress to produce a highly active transcription factor. *Cell* **107**, 881–891 (2001).
4. Carrasco, D. R. *et al.* The differentiation and stress response factor XBP-1 drives multiple myeloma pathogenesis. *Cancer Cell* **11**, 349–360 (2007).
5. De Raedt, T. *et al.* Exploiting cancer cell vulnerabilities to develop a combination therapy for ras-driven tumors. *Cancer Cell* **20**, 400–413 (2011).
6. Mahoney, D. J. *et al.* Virus-tumor interactome screen reveals ER stress response can reprogram resistant cancers for oncolytic virus-triggered caspase-2 cell death. *Cancer Cell* **20**, 443–456 (2011).
7. Carey, L., Winer, E., Viale, G., Cameron, D. & Gianni, L. Triple-negative breast cancer: disease entity or title of convenience? *Nature Rev. Clin. Oncol.* **7**, 683–692 (2010).
8. Foulkes, W. D., Smith, I. E. & Reis-Filho, J. S. Triple-negative breast cancer. *N. Engl. J. Med.* **363**, 1938–1948 (2010).
9. Montagner, M. *et al.* SHARP1 suppresses breast cancer metastasis by promoting degradation of hypoxia-inducible factors. *Nature* **487**, 380–384 (2012).
10. The Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61–70 (2012).
11. Idowu, M. O. *et al.* CD44⁺/CD24^{−/low} cancer stem/progenitor cells are more abundant in triple-negative invasive breast carcinoma phenotype and are associated with poor outcome. *Hum. Pathol.* **43**, 364–373 (2012).
12. Lin, Y., Zhong, Y., Guan, H., Zhang, X. & Sun, Q. CD44⁺/CD24[−] phenotype contributes to malignant relapse following surgical resection and chemotherapy in patients with invasive ductal carcinoma. *J. Exp. Clin. Cancer Res.* **31**, 59 (2012).
13. Creighton, C. J. *et al.* Residual breast cancers after conventional therapy display mesenchymal as well as tumor-initiating features. *Proc. Natl Acad. Sci. USA* **106**, 13820–13825 (2009).
14. Iliopoulos, D., Hirsch, H. A. & Struhl, K. An epigenetic switch involving NF- κ B, Lin28, Let-7 MicroRNA, and IL6 links inflammation to cell transformation. *Cell* **139**, 693–706 (2009).
15. Iliopoulos, D., Hirsch, H. A., Wang, G. & Struhl, K. Inducible formation of breast cancer stem cells and their dynamic equilibrium with non-stem cancer cells via IL6 secretion. *Proc. Natl Acad. Sci. USA* **108**, 1397–1402 (2011).
16. Schwab, L. P. *et al.* Hypoxia inducible factor-1 α promotes primary tumor growth and tumor-initiating cell activity in breast cancer. *Breast Cancer Res.* **14**, R6 (2012).
17. Conley, S. J. *et al.* Antiangiogenic agents increase breast cancer stem cells via the generation of tumor hypoxia. *Proc. Natl Acad. Sci. USA* **109**, 2784–2789 (2012).
18. Yang, M. H. *et al.* Direct regulation of TWIST by HIF-1 α promotes metastasis. *Nature Cell Biol.* **10**, 295–305 (2008).
19. Wouters, B. G. & Koritzinsky, M. Hypoxia signalling through mTOR and the unfolded protein response in cancer. *Nature Rev. Cancer* **8**, 851–864 (2008).
20. Spiotto, M. T. *et al.* Imaging the unfolded protein response in primary tumors reveals microenvironments with metabolic variations that predict tumor growth. *Cancer Res.* **70**, 78–88 (2011).
21. Rody, A. *et al.* A clinically relevant gene signature in triple negative and basal-like breast cancer. *Breast Cancer Res.* **13**, R97 (2011).
22. Barnard, G. A. A New Test for 2 \times 2 Tables. *Nature* **156**, 177 (1945).
23. Barnard, G. A. Significance Tests for 2 \times 2 Tables. *Biometrika* **34**, 123–138 (1947).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank W. G. Kaelin, A.-H. Lee, F. Martinon, M. N. Wein and X. Li for critical review of the manuscript. We are grateful to A. L. Richardson, H. Xu and J. Wang for advice and discussions. We thank L. A. Paskett, X. Liu, R. Kim and Y. Liu for technical support. This work was supported by the National Institutes of Health (CA112663 and AI32412 to L.H.G.; R01HG004069 to X.S.L.; K99CA175290 to Y.C.), the Leukemia and Lymphoma Society (to X.C.) and the National Natural Science Foundation of China (NSFC31329003 to X.S.L.).

Author Contributions X.C. and L.H.G. designed the research; X.C., D.I., Q.Z., M.B.G., M.H., E.L., D.Z.H., S.A., B.H., C.T. and M.S. did the experiments; Q.T. and Y.C. performed the bioinformatics analysis; X.S.L. supervised the bioinformatics analysis; M.N., W.L.T., M.B. and S.J.S. contributed to discussions and critical reagents; J.C.C., M.F., M.D.L., H.S. and J.M. contributed to the patient-derived xenograft experiments; X.C. and L.H.G. wrote the paper.

Author Information ChIP-seq and gene expression microarray data have been deposited in the NCBI Gene Expression Omnibus and are accessible through GEO series accession number GSE49955. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to L.H.G. (lgilme@med.cornell.edu).

Metabolic determinants of cancer cell sensitivity to glucose limitation and biguanides

Kıvanç Birsoy^{1,2,3,4*}, Richard Possemato^{1,2,3,4*}, Franziska K. Lorbeer¹, Erol C. Bayraktar¹, Prathapan Thiru¹, Burcu Yucel¹, Tim Wang^{1,2,3,4}, Walter W. Chen^{1,2,3,4}, Clary B. Clish³ & David M. Sabatini^{1,2,3,4}

As the concentrations of highly consumed nutrients, particularly glucose, are generally lower in tumours than in normal tissues^{1,2}, cancer cells must adapt their metabolism to the tumour microenvironment. A better understanding of these adaptations might reveal cancer cell liabilities that can be exploited for therapeutic benefit. Here we developed a continuous-flow culture apparatus (Nutrostat) for maintaining proliferating cells in low-nutrient media for long periods of time, and used it to undertake competitive proliferation assays on a pooled collection of barcoded cancer cell lines cultured in low-glucose conditions. Sensitivity to low glucose varies amongst cell lines, and an RNA interference (RNAi) screen pinpointed mitochondrial oxidative phosphorylation (OXPHOS) as the major pathway required for optimal proliferation in low glucose. We found that cell lines most sensitive to low glucose are defective in the OXPHOS upregulation that is normally caused by glucose limitation as a result of either mitochondrial DNA (mtDNA) mutations in complex I genes or impaired glucose utilization. These defects predict sensitivity to biguanides, antidiabetic drugs that inhibit OXPHOS^{3,4}, when cancer cells are grown in low glucose or as tumour xenografts. Notably, the biguanide sensitivity of cancer cells with mtDNA mutations was reversed by ectopic expression of yeast NDI1, a ubiquinone oxidoreductase that allows bypass of complex I function⁵. Thus, we conclude that mtDNA mutations and impaired glucose utilization are potential biomarkers for identifying tumours with increased sensitivity to OXPHOS inhibitors.

As nutrient concentrations in tumours are different to those in normal tissues, cancer cells *in vivo* may have metabolic dependencies that are not shared by normal cells⁶. In particular, tumour glucose concentrations are frequently 3- to 10-fold lower than in non-transformed tissues^{1,7}, probably as a result of the high rate of glucose consumption by cancer cells and the poor tumour vasculature. To study the metabolic dependencies imposed on cancer cells by a chronically low-glucose environment, we developed a continuous-flow culture system for maintaining proliferating cells in reduced but steady glucose concentrations for long periods of time. In this system, which we call a Nutrostat, media of a defined glucose concentration is fed into a suspension culture while spent media is removed at the same rate (Fig. 1a). By measuring cell proliferation and glucose concentrations, glucose consumption can be predicted and glucose levels in the intake media can be adjusted so that culture glucose concentrations remain within a 0.5-mM window (Fig. 1b). Jurkat leukaemia cells seeded into 1 mM glucose media in a traditional culture vessel rapidly ceased proliferating as glucose became depleted (Extended Data Fig. 2). In contrast, in a Nutrostat maintained at approximately 0.75 mM glucose, Jurkat cells proliferated exponentially at a rate that was only slightly less than in approximately 10 mM glucose (doubling time of 26 versus 24 h, Fig. 1b). Despite having a small effect on Jurkat cell proliferation, long-term culture in low glucose caused profound metabolic changes: rates of glucose consumption, lactate production

and ATP levels decreased as did levels of intermediates in the upper glycolysis and pentose-phosphate pathways (Fig. 1c, d).

To determine whether all cancer cells respond similarly to long-term low-glucose culture, we undertook a competitive proliferation assay with a pooled collection of 28 patient-derived cancer cell lines, each marked with a lentivirally transduced DNA barcode (Fig. 2a). All cell lines were capable of proliferating in suspension and many were derived from blood cancers but also from breast, lung, stomach and colon cancers. The relative abundance of each cell line at the initial seeding and after three weeks in culture at 0.75 or 10 mM glucose was determined by

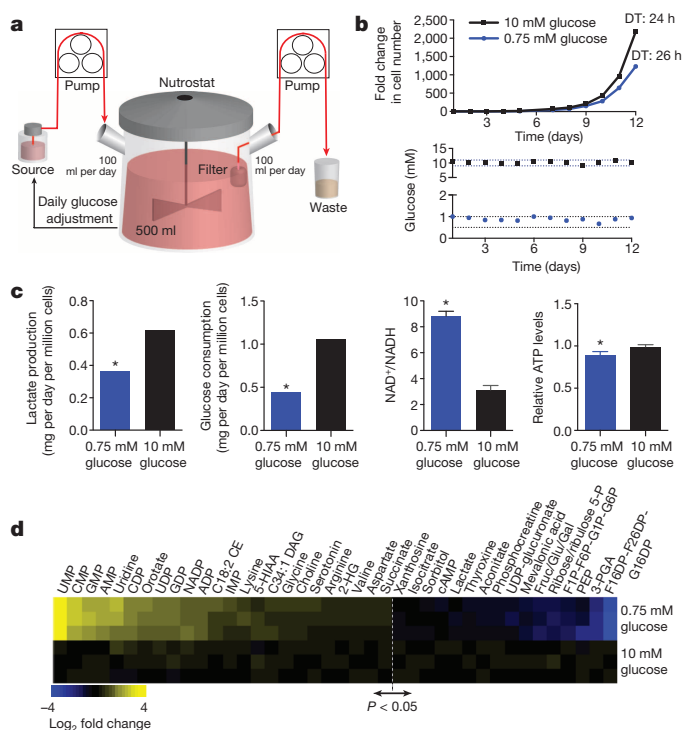


Figure 1 | Nutrostat design and metabolic characterization of cancer cells under chronic glucose limitation. **a**, Nutrostat schematic. **b**, Fold change in cell number (top) and media glucose concentration (bottom) of Jurkat cells grown in Nutrostats at 10 mM (black) or 0.75 mM (blue) glucose. DT, doubling time. **c**, Indicated metabolite levels in Nutrostats at 10 mM (black) or 0.75 mM (blue) glucose. **d**, Differential intracellular metabolite abundances ($P < 0.05$) from cells in Nutrostats at 10 mM (bottom three rows) or 0.75 mM (top three rows) glucose, relative to the average abundance in 10 mM glucose. Colour bar indicates scale (\log_2 transformed). Error bars are s.e.m. ($n = 2$ (glucose and lactate), 3 (NAD(H) ratio) and 8 for ATP levels). Replicates are biological, means reported. * $P < 0.05$ by two-sided Student's *t*-test. C18:2 and C34:1, the number of carbons and number of unsaturated linkages (separated by colons).

¹Whitehead Institute for Biomedical Research, Nine Cambridge Center, Cambridge, Massachusetts 02142, USA. ²Howard Hughes Medical Institute and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ³Broad Institute of Harvard and MIT, Seven Cambridge Center, Cambridge, Massachusetts 02142, USA. ⁴The David H. Koch Institute for Integrative Cancer Research at MIT, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA.

*These authors contributed equally to this work.

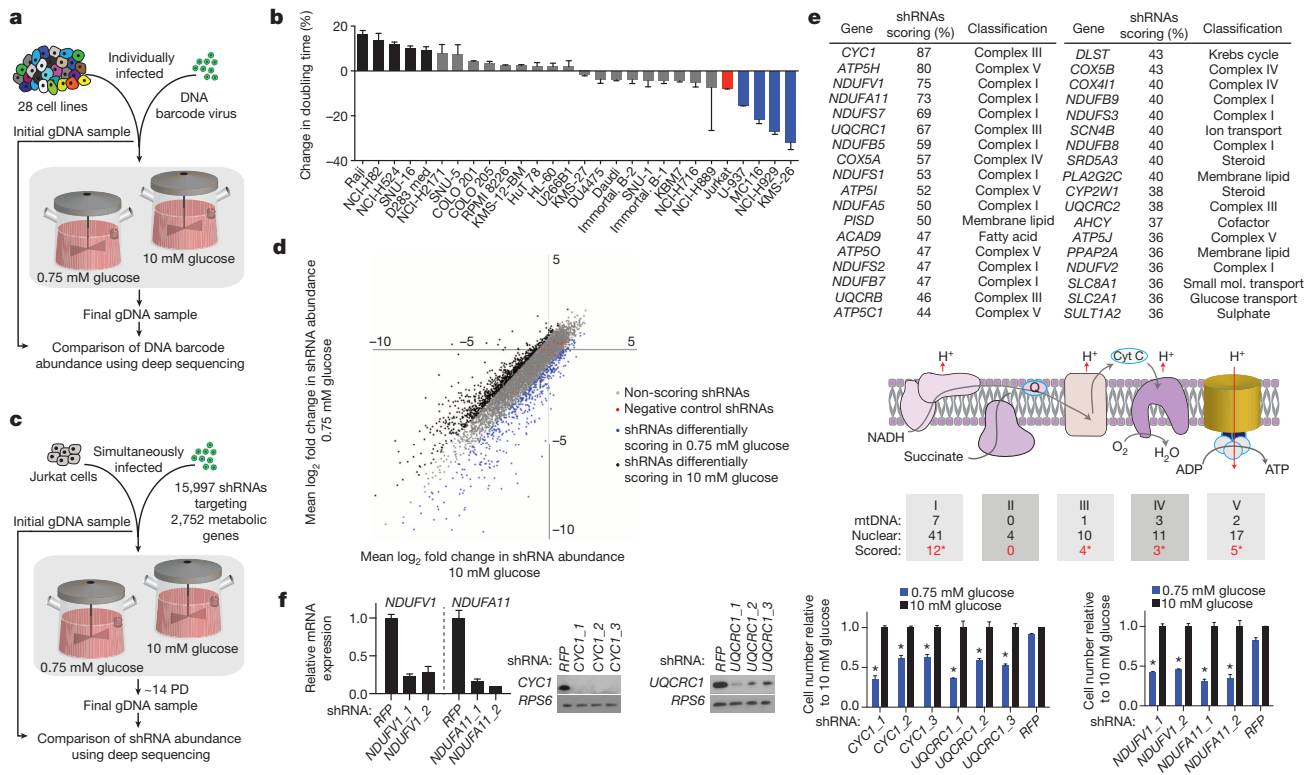


Figure 2 | Barcode-based cell competition assay and RNAi screen in Nutrostats. **a**, Experimental design of cell competition assay. gDNA, genomic DNA. **b**, Per cent changes in doubling times of indicated cell lines in the competition assay, benchmarked to Jurkat cells (red), relative to 0.75 mM glucose. Significant increase (black) or decrease (blue) in doubling time indicated ($P < 0.05$). **c**, Experimental design outline of RNAi-based screen. PD, population doubling. **d**, Primary screening data (mean \log_2 fold change) in 10 mM (x axis) versus 0.75 mM (y axis) glucose. **e**, Genes scoring as preferentially required in 0.75 mM glucose (top); diagram of mitochondrial

OXPHOS Complexes (middle); and the number of mitochondria- or nuclear-encoded components and number of nuclear-encoded genes that scored in **e** are indicated (bottom, red). *P* values were obtained by chi-squared tests for complex I (* $P < 9.3 \times 10^{-49}$), III (* $P < 6.6 \times 10^{-20}$), IV ($P < 8.3 \times 10^{-10}$) and V ($P < 5.6 \times 10^{-19}$). Q, ubiquinone. **f**, Gene suppression of cells expressing indicated shRNAs (left) and proliferation (right) in 0.75 mM (blue) relative to 10 mM glucose (black). * $P < 0.05$ relative to control shRNA (shRFP) by two-sided Student's *t*-test, 0.75 mM glucose. Error bars are s.e.m. ($n = 3$). Replicates are biological, means reported.

deep sequencing of the barcodes, and the change in doubling time calculated for each cell line (Fig. 2b and Supplementary Table 1). Interestingly, cancer cell lines exhibit diverse responses to glucose limitation, as the proliferation of many was unaffected, whereas that of a subset was strongly reduced and another, unexpectedly, increased (Fig. 2b). The presence or absence of known oncogenic mutations did not correlate with differences in low-glucose sensitivity.

To understand the metabolic processes that mediate the response to glucose limitation, we used a cell line of modest glucose sensitivity (Jurkat) to undertake a pooled RNAi screen of 2,752 human metabolic enzymes and small molecule transporters (15,997 total short hairpin RNAs (shRNAs); 5 to 10 shRNAs per gene) in high- or low-glucose media in Nutrostats (Fig. 2c). For control shRNAs and the great majority of gene-targeting shRNAs, the average fold change in shRNA abundance was similar in both conditions (Fig. 2d and Supplementary Table 2). However, 10.5% of shRNAs were differentially depleted and, based on our hit criteria (see Methods; for alternative criteria⁸ see Supplementary Table 3), we identified 28 and 36 genes whose suppression preferentially inhibited cell proliferation in high or low glucose, respectively (Fig. 2e and Extended Data Fig. 3a). Genes selectively required in 10 mM glucose fell into several pathways but were enriched for glycolytic genes (*GAPDH*, *ALDOA*, *PKM*, *ENO1*; $P < 8.6 \times 10^{-7}$). In marked contrast, genes selectively required under 0.75 mM glucose consisted almost exclusively of the nuclear-encoded components of mitochondrial OXPHOS (Fig. 2e). Among the scoring genes were six of the seven nuclear-encoded core complex I subunits conserved between mammals and bacteria⁹, a significant enrichment compared to non-core subunits ($P < 0.0012$, Extended Data Fig. 3b). Two genes required for OXPHOS function,

ACAD9 and *PISD*^{9,10}, also scored, as did *SLC2A1*, the gene encoding the GLUT1 glucose transporter. Short-term individual assays validated that efficient suppression of top scoring OXPHOS genes selectively decreased proliferation under low glucose, whereas hairpins targeting non-scoring OXPHOS genes did so to a significantly lesser extent (Fig. 2f and Extended Data Fig. 3c). Thus, a screen of metabolic genes pinpointed OXPHOS as the key metabolic process required for optimal proliferation of cancer cells under glucose limitation.

Given these very clear results it seemed likely that fundamental differences in mitochondrial function exist between cancer cell lines that are most sensitive (U-937, MC116, NCI-H929, KMS-26) and resistant (Raji, NCI-H82, NCI-H524, SNU-16, NCI-H2171) to glucose limitation. However, we did not detect any differences in mtDNA content, mitochondrial mass, and baseline oxygen consumption rate (OCR, Extended Data Fig. 4a, b). Therefore, we considered whether a key difference is in the mitochondrial response to glucose limitation, which reversibly induces oxygen consumption¹¹ (Extended Data Fig. 4c). Indeed, when cultured in low-glucose media, the low-glucose-sensitive cell lines upregulated OCR less than the resistant ones (Fig. 3a). Furthermore, in response to the mitochondrial uncoupling agent FCCP (carbonyl cyanide *p*-trifluoromethoxyphenylhydrazone), low-glucose-sensitive cell lines induced OCR to a lesser extent than resistant lines, indicating they have reduced spare respiratory capacities (Fig. 3b).

We considered two explanations for why low-glucose-sensitive lines do not substantially increase oxygen consumption upon glucose limitation: first, a defect in glucose utilization that limits substrates for mitochondria; or second, a defect in OXPHOS itself. Consistent with the first possibility, glucose consumption (Fig. 3c) and import (Extended Data

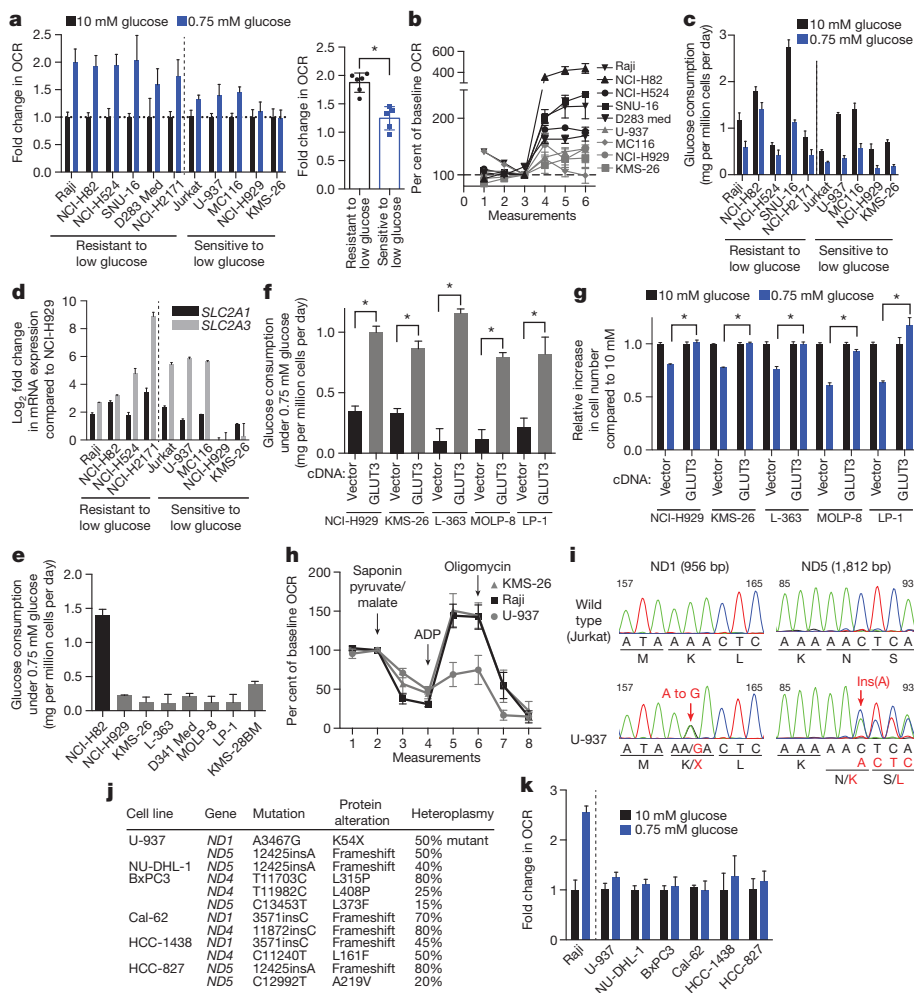


Figure 3 | Deficiencies in glucose utilization or complex I underlie low-glucose sensitivity of cancer cells. **a**, Fold change in oxygen consumption rate (OCR) in 0.75 (blue) relative to 10 mM glucose (black) in indicated cell lines individually (left) or in aggregate (right). **b**, Per cent of baseline OCR relative to third measurement and after addition of FCCP (measurements 4–6) in low-glucose-resistant (black) or low-glucose-sensitive (grey) cell lines. **c**, Glucose consumption rate in 10 mM (black) or 0.75 mM glucose (blue). **d**, Expression (qPCR) of *SLC2A1* (black) or *SLC2A3* (grey) in indicated cell lines (log₂ scale relative to NCI-H929). **e**, **f**, Glucose consumption rate of indicated cell lines under 0.75 mM glucose. **g**, Proliferation (4 days) of control (Vector) or GLUT3 overexpressing (GLUT3) cell lines in 10 mM (black) or 0.75 mM glucose (blue). **h**, Per cent of baseline OCR relative to second measurement of saponin-permeabilized lines given indicated substrates. **i**, Sanger sequencing of *ND1* and *ND5* with corresponding wild-type (black) and mutant (red) nucleotide and protein sequences. **j**, mtDNA mutations in complex I genes identified in indicated cell lines. **k**, Fold increase in OCR of indicated cell lines in 0.75 mM (blue) relative to 10 mM glucose (black). Error bars are s.e.m. $n = 6$ for **a**, **b**, **h**, **k**; $n = 5$ for **c**, **e**, **f**; $n = 3$ for **d**, **g**. Replicates are biological, means reported. $*P < 0.05$ by two-sided student's *t*-test.

Fig. 4d) under low-glucose conditions were defective in two (NCI-H929 and KMS-26) of the low-glucose-sensitive cell lines. Analysis of publicly available gene expression data revealed that these cell lines have low expression of the GLUT3 and GLUT1 glucose transporters, which we verified by quantitative PCR (qPCR) (Fig. 3d), as well as lower levels of several glycolytic enzymes (Extended Data Fig. 4e). Using gene expression data for 967 cell lines¹² we identified additional lines with this expression signature and obtained five of them (Extended Data Fig. 5 and Supplementary Table 4). In low-glucose media, the five lines (LP-1, L-363, MOLP-8, D341 Med, KMS-28BM) had the predicted defect in glucose consumption and proliferation, like NCI-H929 and KMS-26 cells (Fig. 3e and Extended Data Fig. 8b). In all cell lines tested (KMS-26, NCI-H929, L-363, LP-1, MOLP-8), GLUT3 overexpression was sufficient to rescue these phenotypes (Fig. 3f, g, and Extended Data Fig. 4f), without substantially affecting proliferation in high glucose (Extended Data Fig. 4g), arguing that a glucose utilization defect can account for why the proliferation of certain cancer cells is sensitive to low glucose. Indeed, in a competitive proliferation assay, overexpression of GLUT3 provides a growth advantage to KMS-26 cells compared to vector infected controls grown under 0.75 to 2.0 mM glucose in culture and in tumour xenografts (Extended Data Fig. 6).

To investigate whether defects in OXPHOS may be a distinct mechanism underlying low-glucose sensitivity, we focused on U-937 cells as they have low basal OCR but normal glucose utilization (Fig. 3c and Extended Data Fig. 4a). Indeed, in permeabilized-cell mitochondrial-function assays, U-937 cells had a profound defect in using substrates for complexes I (pyruvate and malate) and II (succinate), but not complex IV (TMPD (tetramethylphenylenediamine) and ascorbate) (Fig. 3h and Extended Data Fig. 4h). Sequencing of the seven complex I subunits

encoded by the mitochondrial genome (mtDNA) revealed heteroplasmic truncating mutations in *ND1* and *ND5* in U-937 cells (Fig. 3i).

We used available cancer genome resequencing data and information from the literature^{12,13} to identify additional cell lines with mtDNA mutations in complex I subunits and obtained five, including two with the same ND5 mutation as U-937 cells (Fig. 3j and Extended Data Fig. 7). Like U-937, the additional lines (BxPC3, Cal-62, HCC-1438, HCC-827, NU-DHL-1) weakly boosted OCR in low-glucose media (Fig. 3k) and had a proliferation defect in this condition (Extended Data Fig. 8b). To investigate whether these phenotypes are caused by complex I dysfunction, we expressed the *Saccharomyces cerevisiae* *ND11* gene, which catalyses electron transfer from NADH to ubiquinone without proton translocation^{5,14}. *ND11* expression significantly increased the basal OCR of the complex I defective cells (Cal-62, HCC-827, BxPC3, U-937) and partly rescued their proliferation defect in low glucose, but did not substantially affect proliferation in high glucose (Extended Data Fig. 4f, i–l). In an alternative approach, culture of Cal-62 cells for 1.5 months in the presence of a complex I inhibitor (phenformin) produced a population of cells with significantly enriched wild-type mtDNA content and a corresponding decrease in sensitivity to low glucose, changes not observed in cells expressing *ND11* (Extended Data Fig. 9). Taken together, these data identify defective glucose utilization and mitochondrial dysfunction as two distinct mechanisms for conferring sensitivity to glucose limitation on cancer cell lines.

RNAi-mediated suppression of OXPHOS exacerbated the modest sensitivity of Jurkat cells to low glucose (Fig. 2). Thus, we wondered whether low-glucose-sensitive cell lines may be, like LKB1-deficient cells¹⁵, particularly sensitive to the biguanide class of pharmacological OXPHOS inhibitors^{3,4} that includes metformin and the more potent biguanide

phenformin³. Indeed, in low-glucose media, cell lines with mtDNA-encoded complex I mutations (U-937, BxPC3, Cal-62, HCC-1438, HCC-827, NU-DHL-1) or impaired glucose utilization (NCI-H929, KMS-26, LP-1, L-363, MOLP-8, D341 Med, KMS-28BM) were 5- to 20-fold more sensitive to phenformin compared to control cancer cell lines or an immortalized B-cell line (Fig. 4a), and similar results were obtained with metformin or when using direct cell counting as a readout (Extended Data Fig. 8a, b, d). The low-glucose-sensitive cell lines, particularly those with impaired glucose utilization, tended to be more sensitive to phenformin in 0.75 than 10 mM glucose, but substantial sensitivity persisted at 1.5 to 3.0 mM glucose (Fig. 4b and Extended Data Fig. 8c, e). Importantly, in cells with impaired glucose utilization, GLUT3 overexpression almost completely rescued the phenformin sensitivity specific to the low-glucose condition, such that GLUT3-expressing cells in 0.75 mM glucose and control cells in 10 mM glucose were similarly affected by phenformin (Fig. 4c). Likewise, in cells with mutations in complex I, NDI1 expression almost completely rescued the effects of phenformin on proliferation (Fig. 4d) and oxygen consumption (Fig. 4e and Extended Data Fig. 8g). Phenformin sensitivity is restricted to cells with the intermediate levels of mitochondrial dysfunction typically seen in cancer

cells, as cells lacking mtDNA (143B Rho) are insensitive to phenformin but sensitive to low glucose (Extended Data Fig. 8h).

Consistent with these findings and with the low-glucose environment of tumours^{1,2,7}, phenformin inhibited the growth of mouse tumour xenografts derived from cancer cells with mtDNA mutations (Cal-62, U-937) or poor glucose consumption (KMS-26, NCI-H929), but not from cells lacking these defects (NCI-H2171 and NCI-H82) (Fig. 4f, g). The effects of phenformin on tumour xenograft growth were rescued in mtDNA mutant cells by the introduction of NDI1, and in KMS-26 cells by the overexpression of GLUT3 (Fig. 4g and Extended Data Fig. 8f), demonstrating that the effect of phenformin on these xenografts has a cell-autonomous component. Thus, the glucose-utilization gene signature described earlier and mutations in mtDNA-encoded complex I subunits may serve as biomarkers for identifying tumours that are particularly sensitive to phenformin treatment. Such tumours are probably relatively common, as the prevalence of truncating mutations in mtDNA-encoded OXPHOS components is reported to be as high as 16%¹⁶ and we detect the low-glucose-import gene expression signature in at least 5% of cell lines profiled (of which multiple myeloma and small cell lung cancer are significantly enriched). Interestingly, a number of

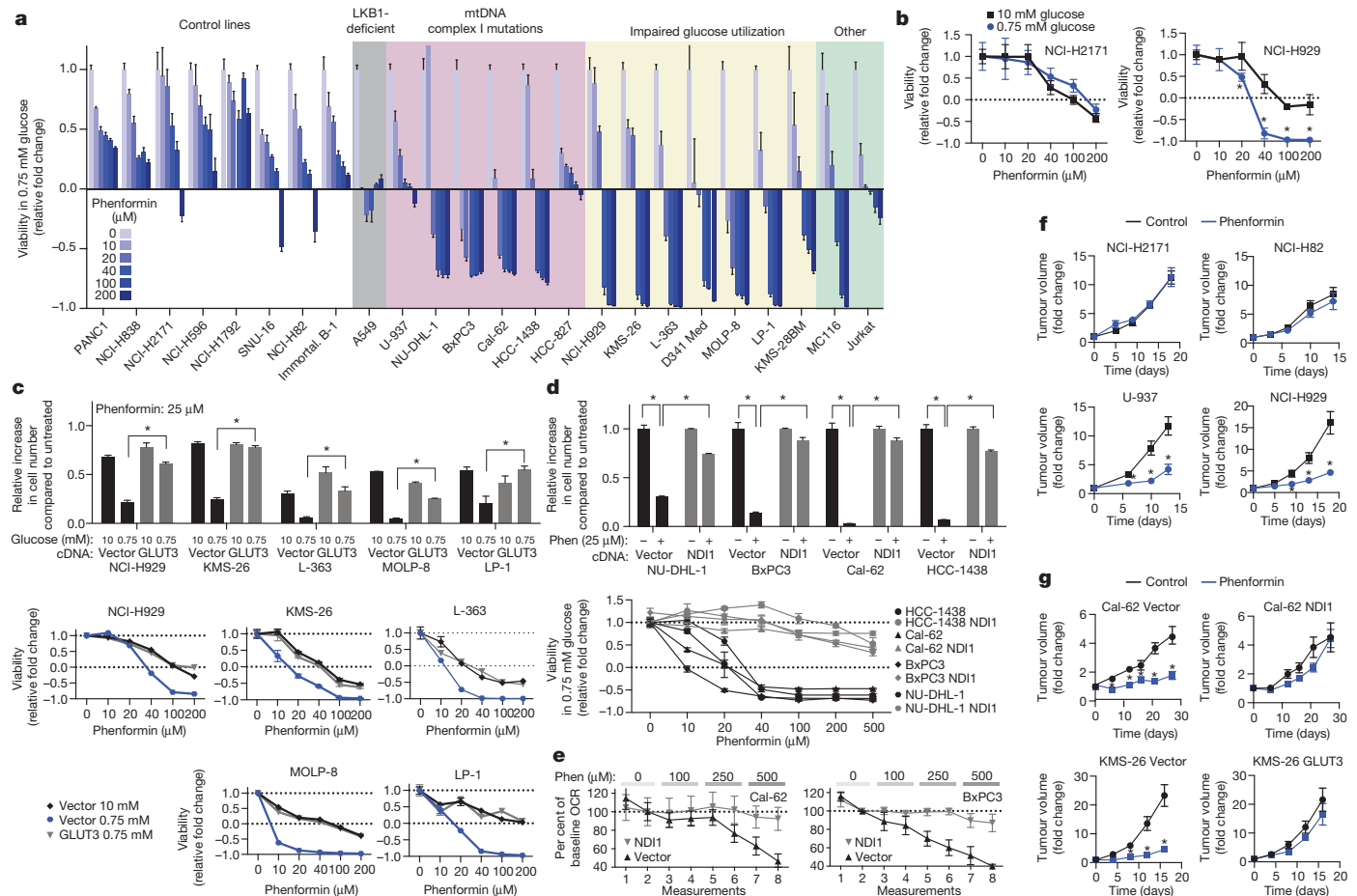


Figure 4 | Cancer cells with deficiencies in glucose utilization or complex I are sensitive to phenformin. **a**, Viability of indicated lines, as measured by ATP levels on day 3 at phenformin concentrations indicated by black–blue scale, in 0.75 mM glucose, compared to ATP levels on day 0. Value of 1 indicates fully viable cells (untreated). Value of 0 indicates no change in ATP level compared to day 0 (cytostatic). Negative values indicate decrease in ATP levels (–1 indicates no ATP). **b**, Viability as in **a** of NCI-H2171 and NCI-H929 cell lines under 0.75 and 10 mM glucose. **c**, Relative increase in cell number (top) and viability as in **a** (bottom) of control (Vector) or GLUT3 overexpressing (GLUT3) cell lines in 10 mM or 0.75 mM glucose at indicated phenformin concentrations relative to untreated cells in 10 mM glucose. **d**, Relative increase in cell number (top) and viability as in **a** (bottom) of vector control (black) or NDI1 (grey) expressing lines in 0.75 mM glucose at indicated phenformin

concentrations relative to untreated cells in 0.75 mM glucose. **e**, Per cent of baseline OCR relative to second measurement of control (Vector) or NDI1-expressing lines (NDI1) relative to the second basal measurement at indicated phenformin concentrations. **f**, Average volume (relative to day 0) of established xenografted tumours derived from control (NCI-H2171, NCI-H82), mtDNA complex I mutant (U-937), or impaired glucose utilization (NCI-H929) cell lines in mice treated with vehicle (black) or phenformin (blue) in drinking water starting at day 0. **g**, Average tumour volume as in **f** of indicated cell lines infected with control, NDI1- or GLUT3-expressing vectors. Error bars are s.e.m. $n = 5$ for **a**, **b**, **c** (bottom), **d** (bottom), **e**, **f**; $n = 6$ (control) or $n = 8$ (GLUT3 or NDI1) for **g**; $n = 3$ for **c** (top), **d** (top). Replicates are biological, means reported. $*P < 0.05$ by two-sided student's t -test.

retrospective studies report survival benefits for cancer patients taking metformin^{17,18}, although it has been unclear whether these effects of metformin are cancer cell-autonomous^{15,19,20}.

In conclusion, we find that cancer cells exhibit diverse responses to glucose limitation and identify defects in glucose utilization and mitochondrial function as major determinants of low-glucose sensitivity (Extended Data Fig. 1). These biomarkers may pinpoint cancer cells that are likely to respond to OXPHOS inhibition alone under tumour-relevant glucose concentrations. Such a targeted strategy may be better tolerated than previously proposed approaches of combining inhibition of OXPHOS and glycolysis^{21–23}. Moreover, our findings underscore the importance of considering glucose concentrations when evaluating the sensitivity of cancer cells to biguanides or other OXPHOS inhibitors. The methods described here should be valuable for studying the responses of cancer cells to tumour-relevant concentrations of other highly consumed nutrients, such as amino acids²⁴, and to additional compounds that target metabolism.

METHODS SUMMARY

Glucose concentrations were maintained in Nutrostats by continuous flow of fresh media at defined glucose concentrations and simultaneous removal of spent media at a fixed rate. Competition assays were carried out at 10 or 0.75 mM glucose in Nutrostats using cell lines barcoded with lentivirally transduced seven-base-pair sequences. Nutrostats were used under similar conditions for the pooled RNAi screen of 2,752 metabolic genes and metabolite transporters. Barcode or shRNA abundances in cells were determined using deep sequencing and compared to initial cultures. Genome resequencing data were used to identify cancer cell lines with mitochondrial genome mutations and the Seahorse Extracellular Flux Analyzer was used to assay mitochondrial function. Defects in glucose import were identified using expression analyses of genes encoding glucose transporters and glycolytic enzymes.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 8 August 2013; accepted 29 January 2014.

Published online 16 March 2014.

- Hirayama, A. *et al.* Quantitative metabolome profiling of colon and stomach cancer microenvironment by capillary electrophoresis time-of-flight mass spectrometry. *Cancer Res.* **69**, 4918–4925 (2009).
- Gullino, P. M., Grantham, F. H. & Courtney, A. H. Glucose consumption by transplanted tumors *in vivo*. *Cancer Res.* **27**, 1031–1040 (1967).
- Owen, M. R., Doran, E. & Halestrap, A. P. Evidence that metformin exerts its anti-diabetic effects through inhibition of complex I of the mitochondrial respiratory chain. *Biochem. J.* **348**, 607–614 (2000).
- El-Mir, M. Y. *et al.* Dimethylbiguanide inhibits cell respiration via an indirect effect targeted on the respiratory chain complex I. *J. Biol. Chem.* **275**, 223–228 (2000).
- Seo, B. B., Matsuno-Yagi, A. & Yagi, T. Modulation of oxidative phosphorylation of human kidney 293 cells by transfection with the internal rotenone-insensitive NADH-quinone oxidoreductase (ND1) gene of *Saccharomyces cerevisiae*. *Biochim. Biophys. Acta* **1412**, 56–65 (1999).
- Cairns, R. A., Harris, I. S. & Mak, T. W. Regulation of cancer cell metabolism. *Nature Rev. Cancer* **11**, 85–95 (2011).
- Urasaki, Y., Heath, L. & Xu, C. W. Coupling of glucose deprivation with impaired histone H2B monoubiquitination in tumors. *PLoS ONE* **7**, e36775 (2012).
- Luo, B. *et al.* Highly parallel identification of essential genes in cancer cells. *Proc. Natl Acad. Sci. USA* **105**, 20380–20385 (2008).

- Tasseva, G. *et al.* Phosphatidylethanolamine deficiency in Mammalian mitochondria impairs oxidative phosphorylation and alters mitochondrial morphology. *J. Biol. Chem.* **288**, 4158–4173 (2013).
- Haack, T. B. *et al.* Exome sequencing identifies ACAD9 mutations as a cause of complex I deficiency. *Nature Genet.* **42**, 1131–1134 (2010).
- Crabtree, H. G. Observations on the carbohydrate metabolism of tumours. *Biochem. J.* **23**, 536–545 (1929).
- Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–607 (2012).
- Jones, J. B. *et al.* Detection of mitochondrial DNA mutations in pancreatic cancer offers a “mass”-ive advantage over detection of nuclear DNA mutations. *Cancer Res.* **61**, 1299–1304 (2001).
- Santidrian, A. F. *et al.* Mitochondrial complex I activity and NAD⁺/NADH balance regulate breast cancer progression. *J. Clin. Invest.* **123**, 1068–1081 (2013).
- Shackelford, D. B. *et al.* LKB1 inactivation dictates therapeutic response of non-small cell lung cancer to the metabolism drug phenformin. *Cancer Cell* **23**, 143–158 (2013).
- Larman, T. C. *et al.* Spectrum of somatic mitochondrial mutations in five cancers. *Proc. Natl Acad. Sci. USA* **109**, 14087–14091 (2012).
- Evans, J. M., Donnelly, L. A., Emslie-Smith, A. M., Alessi, D. R. & Morris, A. D. Metformin and reduced risk of cancer in diabetic patients. *Br. Med. J.* **330**, 1304–1305 (2005).
- DeCensi, A. *et al.* Metformin and cancer risk in diabetic patients: a systematic review and meta-analysis. *Cancer Prev. Res. (Phila.)* **3**, 1451–1461 (2010).
- Birsoy, K., Sabatini, D. M. & Possemato, R. Untuning the tumor metabolic machine: targeting cancer metabolism: a bedside lesson. *Nature Med.* **18**, 1022–1023 (2012).
- Pollak, M. Metformin and pancreatic cancer: a clue requiring investigation. *Clin. Cancer Res.* **18**, 2723–2725 (2012).
- Hall, A. *et al.* Dysfunctional oxidative phosphorylation makes malignant melanoma cells addicted to glycolysis driven by the (V600E)BRAF oncogene. *Oncotarget* **4**, 584–599 (2013).
- Ben Sahra, I., Tanti, J. F. & Bost, F. The combination of metformin and 2 deoxyglucose inhibits autophagy and induces AMPK-dependent apoptosis in prostate cancer cells. *Autophagy* **6**, 670–671 (2010).
- Javeshghani, S. *et al.* Carbon source and myc expression influence the antiproliferative actions of metformin. *Cancer Res.* **72**, 6257–6267 (2012).
- Jain, M. *et al.* Metabolite profiling identifies a key role for glycine in rapid cancer cell proliferation. *Science* **336**, 1040–1044 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank G. Stephanopoulos for assistance with Nutrostat design, M. Holland for mtDNA sequencing consultation, L. Garraway for assistance identifying mtDNA mutant cell lines, T. Yagi for the NDI1 antibody, T. DiCesare for diagrams, and members of the Sabatini Laboratory for assistance (particularly A. Saucedo, C. Koch, O. Yilmaz, Y. Gultekin and A. Hutchins for technical assistance, and D. Lamming and W. Comb for critical reading of the manuscript). This research is supported by fellowships from The Leukemia and Lymphoma Society and The Jane Coffin Childs Fund to K.B., the Council of Higher Education Turkey and Karadeniz T. University Scholarships to B.Y. and grants from the David H. Koch Institute for Integrative Cancer Research at MIT, The Alexander and Margaret Stewart Trust Fund, the NIH (K99 CA168940 to R.P. and CA103866, CA129105, and AI07389 to D.M.S.) and the Starr Cancer Consortium. D.M.S. is an investigator of the Howard Hughes Medical Institute.

Author Contributions K.B., R.P. and D.M.S. conceived the project and designed the experiments. K.B. and R.P. designed and engineered the Nutrostat and performed the screening, knockdown, cell proliferation, extracellular flux, glucose consumption, and tumour formation experiments and processed and analysed sequencing and expression data. F.K.L., E.C.B., B.Y., and W.W.C. assisted with experiments. C.B.C. performed the metabolite profiling experiments. T.W. provided bioinformatic support for shRNA abundance deconvolution, P.T. assisted in identifying mtDNA mutations. K.B., R.L.P. and D.M.S. wrote and all authors edited the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Readers are welcome to comment on the online version of the paper. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.M.S. (sabatini@wi.mit.edu).

Tumour cell heterogeneity maintained by cooperating subclones in Wnt-driven mammary cancers

Allison S. Cleary^{1,2}, Travis L. Leonard^{1,2}, Shelley A. Gestl^{1,2} & Edward J. Gunther^{1,2,3}

Cancer genome sequencing studies indicate that a single breast cancer typically harbours multiple genetically distinct subclones^{1–4}. As carcinogenesis involves a breakdown in the cell–cell cooperation that normally maintains epithelial tissue architecture, individual subclones within a malignant microenvironment are commonly depicted as self-interested competitors^{5,6}. Alternatively, breast cancer subclones might interact cooperatively to gain a selective growth advantage in some cases. Although interclonal cooperation has been shown to drive tumorigenesis in fruitfly models^{7,8}, definitive evidence for functional cooperation between epithelial tumour cell subclones in mammals is lacking. Here we use mouse models of breast cancer to show that interclonal cooperation can be essential for tumour maintenance. Aberrant expression of the secreted signalling molecule Wnt1 generates mixed-lineage mammary tumours composed of basal and luminal tumour cell subtypes, which purportedly derive from a bipotent malignant progenitor cell residing atop a tumour cell hierarchy⁹. Using somatic *Hras* mutations as clonal markers, we show that some Wnt tumours indeed conform to a hierarchical configuration, but that others unexpectedly harbour genetically distinct basal *Hras* mutant and luminal *Hras* wild-type subclones. Both subclones are required for efficient tumour propagation, which strictly depends on luminally produced Wnt1. When biclonal tumours were challenged with Wnt withdrawal to simulate targeted therapy, analysis of tumour regression and relapse revealed that basal subclones recruit heterologous Wnt-producing cells to restore tumour growth. Alternatively, in the absence of a substitute Wnt source, the original subclones often evolve to rescue Wnt pathway activation and drive relapse, either by restoring cooperation or by switching to a defector strategy. Uncovering similar modes of interclonal cooperation in human cancers may inform efforts aimed at eradicating tumour cell communities.

Cancer progression is known to depend on cooperation between tumour cells and neighbouring host cells in the microenvironment. Some have suggested that cooperation between distinct tumour cell subsets may also contribute to the malignant phenotype^{10–12}. Favouring this possibility, genetically distinct subclones cooperatively enhanced tumour growth in models engineered to recapitulate a form of tumour cell heterogeneity identified in brain cancers¹³. Similarly, phenotypically distinct tumour cell subsets cooperatively enhanced tumour invasion in a murine lung cancer model¹⁴. In the case of human breast cancer, recent studies highlight the phenotypic and genetic diversity present locally within individual tumours^{15,16}, but whether this heterogeneity is a cause or a consequence of tumour progression remains unclear. Accordingly, we sought definitive evidence for functional cooperation between tumour cell subsets in mouse models of human breast cancer.

Mammary cancers arising in the classic mouse mammary tumour virus (MMTV)-Wnt1 transgenic mouse model¹⁷ display tumour cell heterogeneity that is widely attributed to malignant transformation of a bipotent mammary progenitor cell^{9,18,19}. Concordantly, MMTV-Wnt1 tumour cells partition into basal and luminal subsets which comingle, recalling the corresponding basal and luminal lineages found in the normal

mammary gland (Fig. 1a, b). Although mutations in Wnt pathway components are rare in human breast cancers, the transcriptional profile of Wnt1-initiated tumours resembles that of other mammary cancer models that commonly show mixed-lineage histopathology, including chemical carcinogen-induced rodent mammary cancers^{20,21}.

While studying cooperating oncogenic mutations in the MMTV-Wnt1 model, we found evidence suggesting that some Wnt tumours harbour unexpected genetic heterogeneity. About half of all Wnt-initiated mammary tumours spontaneously acquire somatic *Hras* mutations that encode an activated oncoprotein^{22,23}. Because *Hras* mutations act dominantly, *Hras* mutant allele fractions (MAFs) of approximately 0.5 are expected, barring copy number changes at the *Hras* locus. Instead, when tumour-derived *Hras* alleles were amplified by PCR and subjected to DNA sequencing, chromatogram peak heights often indicated smaller *Hras* MAFs, with fractions <0.3 detected in four out of ten tumours. Notably, tumours maintained their small *Hras* MAFs as a stable property when explanted onto the flanks of syngeneic host mice (Fig. 1c). This discrepancy could not be explained by contamination of samples with normal (non-tumour) cells as tumour cell content assessed by histopathology consistently exceeded 80%. Moreover, copy number variations leading to either *Hras* wild-type (*Hras*^{wt}) allele gain or *Hras* mutant (*Hras*^{mut}) allele loss seemed unlikely driver events. Instead, we considered whether some Wnt tumours might harbour distinct *Hras*^{mut} and *Hras*^{wt} subclones, noting that biclonal tumours would adopt a mixed-lineage phenotype provided each subclone were committed to a distinct lineage.

To search for lineage-restricted *Hras*^{mut} and *Hras*^{wt} subclones, dissociated cells prepared from *Hras*^{mut} Wnt tumours were sorted into basal and luminal subsets (Extended Data Fig. 1), then *Hras* MAFs were determined for each subset and for corresponding samples of unsorted cells. Half of the *Hras*^{mut} Wnt tumours analysed (five out of ten) showed negligible subset-specific enrichment in *Hras*^{mut} alleles, a pattern consistent with a hierarchical configuration (Fig. 1d, e). In these cases, basal and luminal cells from the same tumour always harboured identical *Hras*^{mut} alleles (Fig. 1e), suggesting that they descended from a common bipotent *Hras*^{mut} progenitor. By contrast, for the remaining half of tumours analysed, *Hras*^{mut} alleles were highly enriched within the basal tumour cell subset, a pattern consistent with a biclonal configuration (Fig. 1e). Basal *Hras*^{mut} allele enrichment correlated with a lower overall *Hras* MAF, further suggesting the presence of a private, subclone-restricted mutation. Regardless of whether the distribution of *Hras*^{mut} alleles fits a hierarchical or biclonal pattern, tumours showed classic mixed-lineage histopathology (Extended Data Fig. 2), and luminal tumour cells were invariably the main source of Wnt1 expression as reported previously²⁴ (Fig. 1f). Therefore, some Wnt tumours appeared to harbour distinct basal *Hras*^{mut} Wnt1^{low} and luminal *Hras*^{wt} Wnt1^{high} subclones, implicating interclonal cooperation in tumour maintenance. These findings recall early reports in which MMTV-associated mammary tumours initiated by activation of endogenous Wnt genes were sometimes noted to be oligoclonal^{25,26}.

¹Jake Gittlen Laboratories for Cancer Research, Pennsylvania State University College of Medicine, Hershey, Pennsylvania 17033, USA. ²Penn State Hershey Cancer Institute, Pennsylvania State University College of Medicine, Hershey, Pennsylvania 17033, USA. ³Department of Medicine (Hematology/Oncology), Pennsylvania State University College of Medicine, Hershey, Pennsylvania 17033, USA.

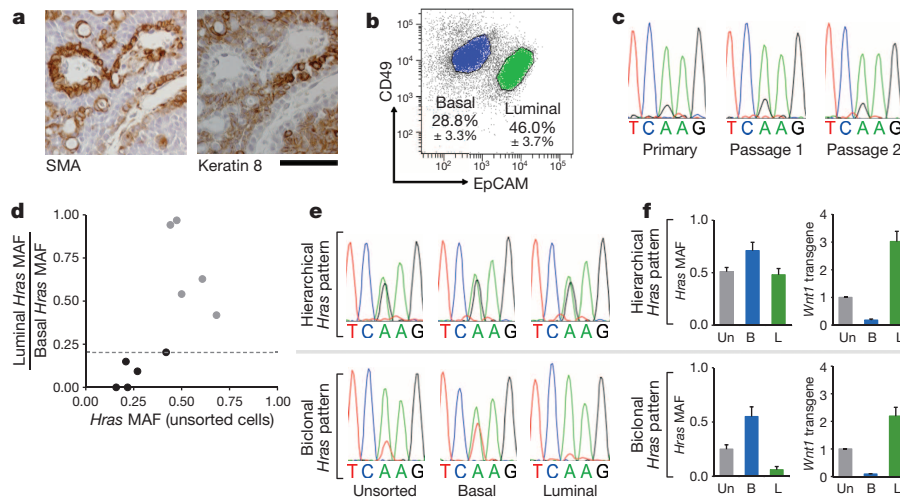


Figure 1 | Evidence for distinct basal *Hras*^{mut} *Wnt1*^{low} and luminal *Hras*^{wt} *Wnt1*^{high} subclones within some MMTV-*Wnt1* tumours.

a, Immunostaining for smooth muscle actin (SMA) and keratin 8 performed on serial sections of a representative MMTV-*Wnt1* mammary tumour. Scale bar, 50 μ m. **b**, Separation of MMTV-*Wnt1* tumour cells into basal (CD49^{high} EpCAM^{low}) and luminal (CD49^{low} EpCAM^{high}) cell subpopulations (where EpCAM denotes epithelial cell adhesion molecule) by flow cytometry. Percentages depict mean \pm s.e.m. for $n = 10$ tumours. **c**, DNA sequencing chromatograms depicting an *Hras*^{CAA61CGA} mutation appearing at a fixed MAF during serial propagation of an MMTV-*Wnt1* tumour. **d**, Graphic depiction of the ratio of luminal to basal *Hras* MAFs plotted against the MAF for unsorted cells. Dotted line depicts the threshold at which

tumours show fivefold *Hras*^{mut} allele enrichment in basal versus luminal cells. Black circles denote tumour values with >5 -fold basal enrichment. **e**, DNA sequencing chromatograms depicting an *Hras*^{CAA61CTA} mutation (top panels) and an *Hras*^{CAA61CGA} mutation (bottom panels) detected in representative *Wnt1* tumours whose basal *Hras*^{mut} allele enrichment fit a hierarchical pattern or biclinal pattern, respectively. **f**, Tumour cell populations analysed by DNA sequencing and by quantitative reverse-transcriptase PCR (qRT-PCR) for *Wnt1* expression relative to *Gapdh*. Histograms on the left show *Hras* MAFs determined from chromatogram peak heights. Histograms on the right show relative *Wnt1* expression with values from unsorted tumour cells set at 1. B, basal; L, luminal; Un, unsorted. Data represent mean \pm s.e.m. for $n = 5$ tumours of each pattern.

Seeking stringent proof that some *Wnt* tumours are biclinal and require interclonal cooperation for maintenance, we attempted to rescue growth of basal *Hras*^{mut} *Wnt1*^{low} subclones from *Wnt1* deprivation by providing access to replacement *Wnt1*-producing cells. For these experiments, the original MMTV-*Wnt1* model (hereafter termed cWnt, denoting constitutive *Wnt1* expression) was used in combination with a closely related model engineered for doxycycline (Dox)-dependent transgene expression (MMTV-rtTA/Tet-O-*Wnt1*; hereafter termed iWnt, denoting inducible *Wnt1* expression). During chronic Dox treatment, iWnt mice and mammary tumours phenocopy their cWnt counterparts, but iWnt tumours regress following Dox withdrawal owing to abrogation of *Wnt1* transgene expression²⁷. To enable tracking of cell lineages in tumour reconstitution experiments, iWnt mice were crossed with a monomeric red fluorescent protein (mRFP) reporter line, generating iWnt/mRFP⁺ mice. As expected, a subset of Dox-dependent iWnt/mRFP⁺ mammary tumours appeared to be biclinal, as they harboured a basally restricted *Hras*^{mut} subclone. After dissociating these tumours into cell suspensions, 10⁵ unsorted cells were injected orthotopically into the mammary fat pads of two sets of Dox-treated, mRFP-reporter-negative female host mice (Fig. 2a). Control hosts lacked a transgene capable of rescuing tumour cells from *Wnt* withdrawal (wt/mRFP⁻), whereas rescue hosts expressed the constitutive *Wnt1* transgene (cWnt/mRFP⁻).

During chronic Dox treatment, both control and rescue hosts developed mammary tumours in most glands injected with iWnt/mRFP⁺ tumour cells (Fig. 2b and Extended Data Table 1). As expected, these reconstituted tumours usually regressed when iWnt transgene expression was switched off via Dox withdrawal. On control hosts, tumour regression was always complete, and mice remained relapse-free during 6 weeks of monitoring. Interestingly, subclinical disease often persisted, as most control hosts subsequently relapsed after Dox retreatment (Fig. 2b). By contrast, on cWnt rescue hosts, most reconstituted tumours only partially regressed, then relapsed spontaneously (Fig. 2b and Extended Data Fig. 3a). On control hosts, primary tumours were reconstituted almost exclusively from donor mRFP⁺ cells, and relapses triggered by Dox retreatment remained mRFP⁺ as expected (Fig. 2c). By contrast, on

rescue hosts primary tumours showed varying degrees of chimaerism due to incorporation of mRFP⁻ (host-derived) luminal cells (Extended Data Fig. 3b), and relapses arising during Dox withdrawal always showed pronounced lineage-restricted chimaerism, resulting in mRFP⁺/basal and mRFP⁻/luminal subpopulations (Fig. 2c). To confirm that donor basal subclones recruit host luminal epithelium to serve as a replacement *Wnt1* source at relapse, we turned to northern hybridization analysis of tumour RNAs. Notably, tumours reconstituted on rescue hosts typically expressed the larger iWnt transgene before Dox withdrawal (pertinent exceptions discussed in Extended Data Fig. 3), then switched to expressing the smaller cWnt transgene at relapse, indicating heterologous rescue (Fig. 2d).

Furthermore, the biclinal configuration evident in parental tumours was maintained in all reconstituted tumours in that basal cells were *Hras*^{mut} *Wnt1*^{low}, whereas luminal cells were *Hras*^{wt} *Wnt1*^{high} (Extended Data Fig. 4). We repeated these rescue experiments twice, beginning each time with an independent, iWnt/mRFP tumour harbouring a distinct, basally restricted *Hras* mutation. In all cases, we observed rescue of donor-derived basal *Hras*^{mut} *Wnt1*^{low} tumour cells by cWnt host-derived luminal *Hras*^{wt} *Wnt1*^{high} cells (Fig. 3a, b and Extended Data Fig. 5). Moreover, the *Hras*^{mut} allele detected in relapsed tumours was always identical to that detected in parental tumours, confirming that basal subclones found at relapse were descended from donor-derived tumour cells and were not novel clones. Examination of tumour sections by fluorescence microscopy revealed pervasive intermingling of basal mRFP⁺ and luminal mRFP⁻ tumour cells within chimaeric relapses, consistent with the prevailing notion that secreted *Wnt* provides a short-range signal to neighbouring cells (Fig. 3c). Prospective analysis of a larger set of independent *Wnt* tumours will be required to precisely estimate the overall fraction of *Wnt* tumours that is biclonally configured. Notably, we cannot yet determine clonal configurations for the half of *Wnt* tumours that lack an *Hras* mutation.

Whereas biclinal iWnt/mRFP⁺ tumours were readily reconstituted from unsorted (fluorescence-activated cell sorting (FACS)-naïve) cells, sorted basal and luminal cells each reconstituted tumours inefficiently when injected into mammary glands (Extended Data Fig. 6a), perhaps

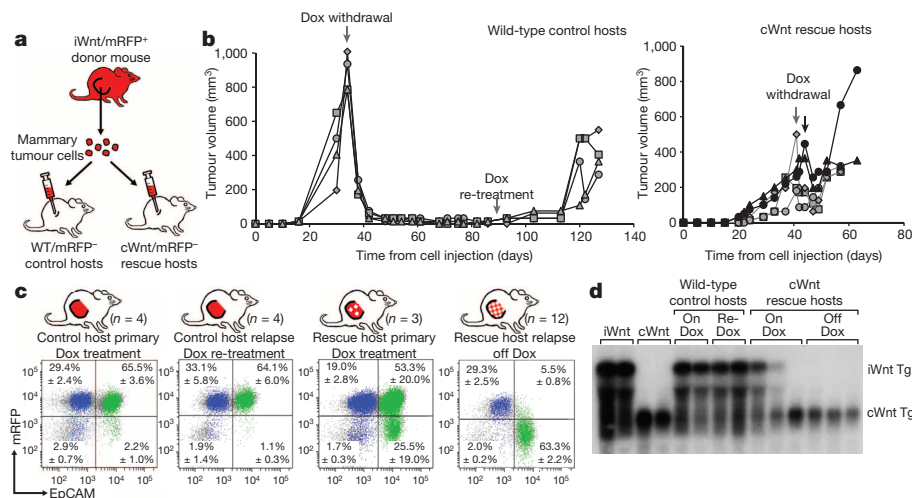


Figure 2 | Rescue of basal *Hras*^{mut} iWnt tumour cells from Wnt withdrawal by heterologous luminal cWnt cells. **a**, Schematic of experimental design. **b**, Growth curves of tumours reconstituted in wild-type or cWnt hosts following injection of iWnt/mRFP⁺ tumour cells. **c**, Representative FACS plots showing contributions by donor-derived mRFP⁺ cells and host-derived mRFP⁻ cells to reconstituted tumours. Percentages depict mean ± s.e.m.

owing in part to loss of cell viability during FACS. Notably, tumours that did arise after injecting a single sorted subtype were always biclonal, comprised of both basal *Hras*^{mut} *Wnt1*^{low} and luminal *Hras*^{wt} *Wnt1*^{high} subsets (seven out of seven tumours analysed, Extended Data Fig. 6b). Given the imperfect separation achieved by FACS (95–98% purity), rare cells cross-contaminating each subset presumably sufficed to permit interclonal cooperation during tumour reconstitution. Consistent with this notion, the relative sizes of the basal and luminal cell populations within these tumours approximated that found in parental tumours and did not reflect the lineage enrichment achieved by sorting. We confirmed this result in an experimental context where the putative cooperating subclones were differentially labelled by the mRFP transgene.

for *n* tumour explants as indicated. Colours indicate events within the basal (blue; CD49^{high} EpCAM^{low}) and luminal (green; CD49^{low} EpCAM^{high}) gates. **d**, Northern hybridization analysis of tumour RNA with *Wnt1* probe. The larger bicistronic iWnt transcript encodes both *Wnt1* and firefly luciferase. Tg, transgene.

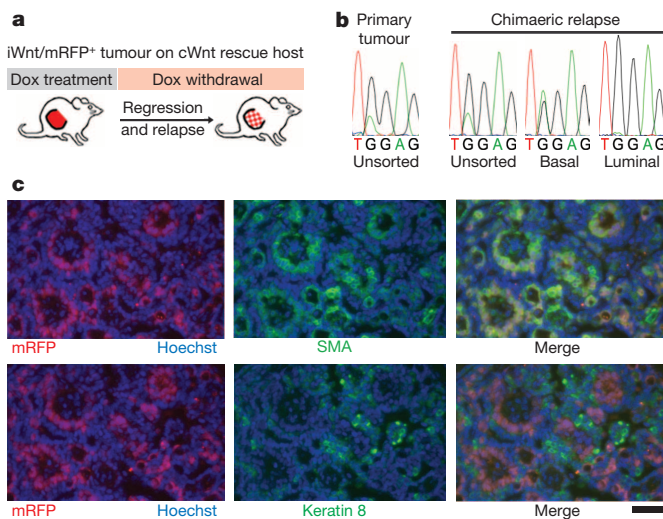


Figure 3 | Lineage-restricted subclones recapitulate mosaic heterogeneity in chimaeric cWnt/iWnt tumours. **a**, Schematic summarizing the experiment culminating in chimaeric tumour relapse. **b**, DNA sequence chromatograms depicting matching *Hras*^{GGA12AGA} mutations detected in unsorted and sorted populations from primary and relapsed tumours as indicated. **c**, Immunostaining of basal (SMA, top panels) and luminal (keratin 8, bottom panels) tumour cells within a Dox-independent relapse arising on a cWnt host. Red fluorescence marks donor-derived iWnt/mRFP⁺ cells intermingled with mRFP⁻ host-derived cells. Scale bar, 50 μm.

Here, tumour cells derived from chimaeric relapses generated in our rescue experiment were studied prospectively. Again, neither the basal (mRFP⁺/Hras^{mut} *Wnt1*^{low}) nor the luminal (mRFP⁻/Hras^{wt} *Wnt1*^{high}) subsets reconstituted tumours efficiently, whereas a 1:1 admixture of both sorted populations reliably reconstituted biclonal tumours (Extended Data Fig. 7). Notably, every tumour reconstituted in these experiments faithfully restored the subclonal composition of the source tumour, pointing to strong selection favouring propagation of both subclones in tandem.

iWnt tumours that regress upon Dox withdrawal frequently relapse weeks later as Dox-independent tumours (DITs), mirroring the clinical scenario of acquired resistance to effective targeted therapy. Next, we asked which subclone(s) contribute when biclonal tumours beget relapse. A putative biclonal iWnt tumour with an *Hras* MAF <0.3 was identified and propagated on host mice, generating a set of Dox-dependent tumour explants. Explants maintained the *Hras* MAF observed in the parental tumour, suggesting a stable biclonal configuration (Fig. 4a, b). Host mice were then subjected to Dox withdrawal and monitored until relapse, generating a set of 20 DITs. In accord with our previous work²⁸, 18 of the 20 relapses (90%) occurred through one of two mutually exclusive modes of Wnt pathway reactivation. Seven DITs (35%) re-expressed the *Wnt1* transgene, and all seven had acquired one of two *rtTA* mutations (G138R or H100Y) previously shown to rescue mammary tumours from oncogene withdrawal by enabling aberrant, Dox-independent expression of TetO-controlled transgenes²⁹. All seven of these tumours had an *Hras* MAF comparable to parental tumour (Fig. 4b), strongly suggesting that *rtTA* mutations originating within the *Hras*^{wt} subclone restored both *Wnt1* expression and cooperation with *Hras*^{mut} cells, culminating in biclonal relapse.

Eleven DITs (55%) instead rescued oncogenic signalling by acquiring an activating mutation in the β -catenin gene (*Ctnnb1*, hereafter *βcat*), a key downstream Wnt effector (Fig. 4a, b). Compared with parental tumour, these relapses showed markedly increased *Hras* MAFs that were highly reproducible across the tumour set (Fig. 4b). Therefore, *βcat* mutations probably originated within *Hras*^{mut} cells that later emerged as predominant relapse clones. By activating the Wnt pathway in a cell-autonomous manner, *βcat* mutations presumably obviated the need to maintain cooperation with Wnt-producing *Hras*^{wt} subclones. As such, *βcat*^{mut} relapse clones act like ‘defectors’ in evolutionary game theory terms. *Hras* MAFs in *βcat*^{mut} relapses consistently exceeded 0.5, indicating that *βcat*^{mut} clones must harbour additional *Hras* locus aberrations; however, no gross changes in *Hras* gene copy number were observed

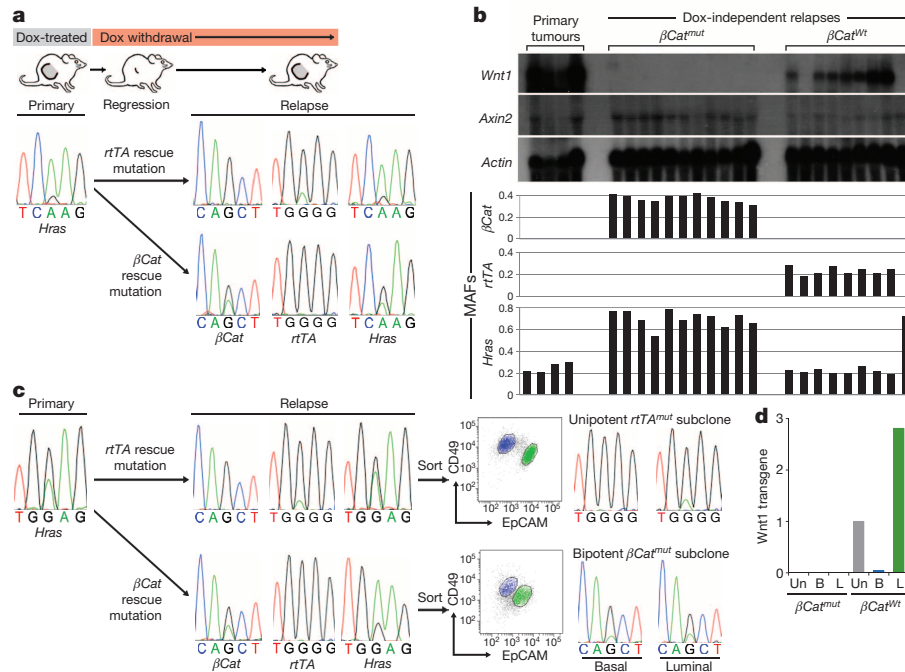


Figure 4 | Relapse of biclonal tumours through the evolution of either subclone. **a**, DNA sequencing chromatograms depicting matching *Hras*^{CAA61CGA} mutations detected in primary and relapsed tumours, with an increased MAF detected in the setting of a β cat mutation. **b**, Bottom, histogram depicting MAFs for a series of primary and relapsed tumours derived from a parental biclonal tumour. Top, corresponding gene expression patterns for each tumour by northern hybridization analysis. **c**, DNA sequencing chromatograms depicting matching *Hras*^{GGA12GAA} mutations detected in

primary and relapsed tumours, with an increased MAF detected in the setting of a β cat mutation. Panels on the right depict analysis of unsorted and sorted cells at relapse showing unipotent or bipotent mutant subclones, depending on the mode of Wnt pathway reactivation. **d**, Histogram showing *Wnt1* expression levels relative to *Gapdh* in unsorted and sorted tumour cells from a β cat^{mut} *rtTA*^{wt} relapse versus a β cat^{wt} *rtTA*^{mut} relapse with the value measured in unsorted cells from the latter relapse set at 1.

(Extended Data Fig. 8), implicating copy-number-neutral loss-of-heterozygosity events.

To further examine the clonal configuration of relapsed tumours, an iWnt/mRFP⁺ tumour previously confirmed as biclonal in our rescue experiments (Fig. 2) was propagated as above to derive DITs, then relapse-derived tumour cells were separated into basal and luminal subsets and analysed. One DIT that relapsed via *Wnt1* transgene re-expression was biclonal with a luminally restricted *rtTA* mutation (Fig. 4c). Trophic support from this luminal *rtTA*^{mut} subclone probably rescued growth of its basal *rtTA*^{wt} counterpart, providing a plausible cellular mechanism whereby this rescue mutation was maintained at a low MAF. By contrast, our previous analysis indicated that β cat rescue mutations originate within basal tumour cells and obviate the need to cooperate with Wnt-producing luminal cells. Nonetheless, β cat^{mut} relapses consistently harboured abundant luminal tumour cells (Fig. 4c and Extended Data Fig. 9). We proposed that acquired β cat mutations endow basally restricted subclones with novel bipotent differentiation potential, thereby converting them to hierarchically configured clones at relapse. Two β cat^{mut} DITs analysed as above showed comparable β cat^{mut} allele prevalence in the basal and luminal subsets (Fig. 4c), consistent with a scenario in which β cat^{mut} relapse clones acquired bipotency. (An alternative scenario in which each subclone independently acquired matching β cat mutations cannot be formally excluded, but seems less likely.) In our previous experiments (Fig. 2 and Extended Data Fig. 4) and in the *rtTA*^{mut} relapse profiled above, this same subclone invariably behaved in a unipotent manner, remaining basally restricted when partnered with a *Wnt1*-expressing luminal subclone in the context of primary and relapsed tumours.

Efforts to explain how some cancers stably maintain intratumoural lineage diversity typically invoke tumour cell hierarchies. Here we show that cooperation between lineage-restricted subclones provides an alternative mechanism for maintaining tumour cell heterogeneity. In our Wnt

models, we found evidence for both hierarchically and biclonally configured tumours, yet differently configured tumours were indistinguishable by histopathology, acquired equivalent cooperating *Hras*^{mut} alleles (albeit with differences in tumour cell compartmentalization), and were comparably *Wnt1* dependent. Thus, although distinct clonal configurations evolved, they converged towards analogous malignant phenotypes. These findings highlight the difficulties associated with inferring the clonal architecture of cancers from histopathology, even in the 'simplified' context of mouse models. Notably, the Wnt models described here provide an experimentally tractable system for exploring whether and how a tumour's clonal configuration determines its clinical behaviour and curability.

Our study does not define when distinct subclones emerge in the course of tumour progression. Interclonal cooperation may be particularly prevalent in tumours initiated by aberrant expression of secreted signalling molecules, such as *Wnt1* and platelet-derived growth factor³⁰. In principle, germline mutations that impart a cancer predisposition also might bias tumours towards a biclonal configuration, as any subsequent cooperation-enabling mutations would necessarily accrue in a cell with mutant neighbours. As such, it will be important to determine whether interclonal cooperation can arise when initiating events originate in somatic cells or act primarily in a cell-intrinsic manner. If cooperation emerges as a common mechanism for maintaining subclone diversity in malignancies, this scenario would counter a key assumption made when interpreting cancer genome sequences. Specifically, certain mutations detected at low allelic fractions and commonly assumed to be late events in tumour progression instead may be early events that enable interclonal cooperation.

METHODS SUMMARY

The MMTV-*Wnt1* (FVB.Cg-Tg(*Wnt1*)1Hev/J; stock no. 002934) and mRFP (B6.Cg-Tg(CAG-mRFP1)1F1Hadj/J; stock no. 005884) transgenic lines were obtained from

the Jackson Laboratories. The MMTV-rtTA and tetO-Wnt1 transgenic lines were a gift from L. Chodosh. All mice either were generated in an inbred FVB/N background or were backcrossed ten or more generations with FVB/N breeders before initiating experiments. To determine *Hras* MAFs, we measured peak height (PH) attributable to mutant and wild-type alleles on DNA sequencing chromatograms using ImageJ 1.46 software, then calculated *Hras* MAF values using the following formula: $MAF = PH^{\text{mutant}} / (PH^{\text{mutant}} + PH^{\text{wild-type}})$. Mammary tumours were dissociated into single-cell suspensions and processed for immunostaining and FACS as previously described²⁴. Tumour cells were re-suspended in a 50% solution of Matrigel in PBS before mammary fat pad injection.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 4 July 2013; accepted 27 February 2014.

- Shah, S. P. *et al.* Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature* **461**, 809–813 (2009).
- Ding, L. *et al.* Genome remodelling in a basal-like breast cancer metastasis and xenograft. *Nature* **464**, 999–1005 (2010).
- Nik-Zainal, S. *et al.* The life history of 21 breast cancers. *Cell* **149**, 994–1007 (2012).
- Navin, N. *et al.* Tumour evolution inferred by single-cell sequencing. *Nature* **472**, 90–94 (2011).
- Nowell, P. C. The clonal evolution of tumor cell populations. *Science* **194**, 23–28 (1976).
- Greaves, M. & Maley, C. C. Clonal evolution in cancer. *Nature* **481**, 306–313 (2012).
- Wu, M., Pastor-Pareja, J. C. & Xu, T. Interaction between Ras^{V12} and scribbled clones induces tumour growth and invasion. *Nature* **463**, 545–548 (2010).
- Ohsawa, S. *et al.* Mitochondrial defect drives non-autonomous tumour progression through Hippo signalling in *Drosophila*. *Nature* **490**, 547–551 (2012).
- Li, Y. *et al.* Evidence that transgenes encoding components of the Wnt signaling pathway preferentially induce mammary cancers from progenitor cells. *Proc. Natl Acad. Sci. USA* **100**, 15853–15858 (2003).
- Heppner, G. H. Cancer cell societies and tumor progression. *Stem Cells* **11**, 199–203 (1993).
- Axelrod, R., Axelrod, D. E. & Pienta, K. J. Evolution of cooperation among tumor cells. *Proc. Natl Acad. Sci. USA* **103**, 13474–13479 (2006).
- Marusyk, A. & Polyak, K. Tumor heterogeneity: causes and consequences. *Biochim. Biophys. Acta* **1805**, 105–117 (2010).
- Inda, M. M. *et al.* Tumor heterogeneity is an active process maintained by a mutant EGFR-induced cytokine circuit in glioblastoma. *Genes Dev.* **24**, 1731–1745 (2010).
- Calbo, J. *et al.* A functional role for tumor cell heterogeneity in a mouse model of small cell lung cancer. *Cancer Cell* **19**, 244–256 (2011).
- Park, S. Y., Gonen, M., Kim, H. J., Michor, F. & Polyak, K. Cellular and genetic diversity in the progression of in situ human breast carcinomas to an invasive phenotype. *J. Clin. Invest.* **120**, 636–644 (2010).
- Navin, N. *et al.* Inferring tumor progression from genomic heterogeneity. *Genome Res.* **20**, 68–80 (2010).
- Tsukamoto, A. S., Grosschedl, R., Guzman, R. C., Parslow, T. & Varmus, H. E. Expression of the *int-1* gene in transgenic mice is associated with mammary gland hyperplasia and adenocarcinomas in male and female mice. *Cell* **55**, 619–625 (1988).
- Liu, B. Y., McDermott, S. P., Khwaja, S. S. & Alexander, C. M. The transforming activity of Wnt effectors correlates with their ability to induce the accumulation of mammary progenitor cells. *Proc. Natl Acad. Sci. USA* **101**, 4158–4163 (2004).
- Cho, R. W. *et al.* Isolation and molecular characterization of cancer stem cells in MMTV-*Wnt-1* murine breast tumors. *Stem Cells* **26**, 364–371 (2008).
- Kim, S., Roopra, A. & Alexander, C. M. A phenotypic mouse model of basaloid breast tumors. *PLoS ONE* **7**, e30979 (2012).
- Herschkowitz, J. I. *et al.* Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome Biol.* **8**, R76 (2007).
- Podsypanina, K., Li, Y. & Varmus, H. E. Evolution of somatic mutations in mammary tumors in transgenic mice is influenced by the inherited genotype. *BMC Med.* **2**, 24 (2004).
- Jang, J. W., Boxer, R. B. & Chodosh, L. A. Isoform-specific ras activation and oncogene dependence during MYC- and Wnt-induced mammary tumorigenesis. *Mol. Cell. Biol.* **26**, 8109–8121 (2006).
- Kim, S., Goel, S. & Alexander, C. M. Differentiation generates paracrine cell pairs that maintain basaloid mouse mammary tumors: proof of concept. *PLoS ONE* **6**, e19310 (2011).
- Mester, J., Wagenaar, E., Sluys, M. & Nusse, R. Activation of *int-1* and *int-2* mammary oncogenes in hormone-dependent and -independent mammary tumors of GR mice. *J. Virol.* **61**, 1073–1078 (1987).
- Roelink, H., Wagenaar, E. & Nusse, R. Amplification and proviral activation of several Wnt genes during progression and clonal variation of mouse mammary tumors. *Oncogene* **7**, 487–492 (1992).
- Gunther, E. J. *et al.* Impact of p53 loss on reversal and recurrence of conditional Wnt-induced tumorigenesis. *Genes Dev.* **17**, 488–501 (2003).
- Debies, M. T. *et al.* Tumor escape in a *Wnt1*-dependent mouse breast cancer model is enabled by p19^{Arf}/p53 pathway lesions but not p16^{Ink4a} loss. *J. Clin. Invest.* **118**, 51–63 (2008).
- Podsypanina, K., Politi, K., Beverly, L. J. & Varmus, H. E. Oncogene cooperation in tumor maintenance and tumor recurrence in mouse mammary tumors induced by *Myc* and mutant *Kras*. *Proc. Natl Acad. Sci. USA* **105**, 5242–5247 (2008).
- Fomchenko, E. I. *et al.* Recruited cells can become transformed and overtake PDGF-induced murine gliomas *in vivo* during tumor progression. *PLoS ONE* **6**, e20605 (2011).

Acknowledgements We thank the benefactors of the Jake Gittlen Laboratories for Cancer Research and the following Penn State College of Medicine staff: L. Budgeon of the Gittlen Histology Core, N. Sheaffer and J. Bednarczyk of the Flow Cytometry Core, D. Stanford of the Molecular Genetics Core, and J. Mohl of the Barrier Rodent Facility. This project was funded with grant support from the National Institutes of Health/ National Cancer Institute, the Mary Kay Foundation, the Donald B. and Dorothy L. Stabler Foundation, and the Pennsylvania Department of Health. A.S.C. is supported by Department of Defense Predoctoral Traineeship Award W81XWH-11-1-0038.

Author Contributions A.S.C. designed and carried out experiments, interpreted data and co-wrote the manuscript. T.L.L. and S.A.G. performed experiments, interpreted data, and provided commentary on the manuscript. E.J.G. designed experiments, interpreted data and co-wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.J.G. (ejg12@psu.edu).

Reversible and adaptive resistance to BRAF(V600E) inhibition in melanoma

Chong Sun^{1*}, Liqin Wang^{1*}, Sidong Huang^{1,2*}, Guus J. J. E. Heynen¹, Anirudh Prahallad¹, Caroline Robert³, John Haanen⁴, Christian Blank⁴, Jelle Wesseling⁵, Stefan M. Willems^{1,6}, Davide Zecchin^{7,8}, Sebastijan Hobor⁸, Prashanth K. Bajpe¹, Cor Liefstink¹, Christina Mateus³, Stephan Vagner³, Wipawadee Grernrum¹, Ingrid Hofland⁵, Andreas Schlicker¹, Lodewyk F. A. Wessels¹, Roderick L. Beijersbergen¹, Alberto Bardelli^{7,8,9}, Federica Di Nicolantonio^{7,8}, Alexander M. M. Eggermont³ & Rene Bernards¹

Treatment of BRAF(V600E) mutant melanoma by small molecule drugs that target the BRAF or MEK kinases can be effective, but resistance develops invariably^{1,2}. In contrast, colon cancers that harbour the same BRAF(V600E) mutation are intrinsically resistant to BRAF inhibitors, due to feedback activation of the epidermal growth factor receptor (EGFR)^{3,4}. Here we show that 6 out of 16 melanoma tumours analysed acquired EGFR expression after the development of resistance to BRAF or MEK inhibitors. Using a chromatin-regulator-focused short hairpin RNA (shRNA) library, we find that suppression of sex determining region Y-box 10 (SOX10) in melanoma causes activation of TGF- β signalling, thus leading to upregulation of EGFR and platelet-derived growth factor receptor- β (PDGFR β), which confer resistance to BRAF and MEK inhibitors. Expression of EGFR in melanoma or treatment with TGF- β results in a slow-growth phenotype with cells displaying hallmarks of oncogene-induced senescence. However, EGFR expression or exposure to TGF- β becomes beneficial for proliferation in the presence of BRAF or MEK inhibitors. In a heterogeneous population of melanoma cells having varying levels of SOX10 suppression, cells with low SOX10 and consequently high EGFR expression are rapidly enriched in the presence of drug, but this is reversed when the drug treatment is discontinued. We find evidence for SOX10 loss and/or activation of TGF- β signalling in 4 of the 6 EGFR-positive drug-resistant melanoma patient samples. Our findings provide a rationale for why some BRAF or MEK inhibitor-resistant melanoma patients may regain sensitivity to these drugs after a 'drug holiday' and identify patients with EGFR-positive melanoma as a group that may benefit from re-treatment after a drug holiday.

Activating mutations in the BRAF oncogene are found in over half of the patients with advanced melanoma^{5,6}. Inhibition of the oncogenic BRAF protein with the small molecule inhibitor PLX4032 (vemurafenib) or its downstream effector MEK with GSK1120212 (trametinib) have shown impressive initial responses in patients with BRAF mutant melanoma^{1,2}. However, single-agent therapies for advanced cancers are rarely curative, due to the rapid development of resistance. To date, several drug resistance mechanisms have been identified in melanomas treated with vemurafenib, including increased expression of the gene encoding the COT kinase, mutation of downstream MEK1 (also known as MAP2K1) kinase, NRAS mutations and amplification or alternative splicing of the BRAF gene^{7–11}. Moreover, increased expression of receptor tyrosine kinases (RTKs) has been observed as a mechanism of BRAF inhibitor resistance^{11–13}.

It has been shown recently that intrinsic resistance of BRAF mutant colon cancers to vemurafenib is the result of feedback activation of

EGFR when BRAF is inhibited^{3,4}. To investigate whether BRAF(V600E) mutant melanoma patients frequently develop resistance to BRAF or MEK inhibitors through acquired expression of EGFR in their tumours, we obtained biopsies from BRAF(V600E) mutant melanomas from 16 patients treated with either the MEK inhibitor trametinib ($n = 1$) or the BRAF inhibitors dabrafenib ($n = 3$) or vemurafenib ($n = 12$). Tumour biopsies collected both before treatment initiation and after the development of drug resistance were stained for EGFR expression. We found that 6 out of 16 post-treatment biopsies gained notable EGFR expression as judged by immunohistochemistry (Fig. 1a, b and Supplementary Table 1).

Melanomas are derived from the neural crest and in general do not express EGFR¹⁴. Hence, acquired EGFR expression during drug selection may represent a stress response that is not favoured in the absence of drug treatment. Indeed, the proliferation rate of A375 BRAF(V600E) melanoma cells³ engineered to express EGFR decreased as the concentration of EGFR ligand increased (Fig. 1c). Moreover, A375 cells that express EGFR also proliferate slowly compared to parental control cells in nude mouse xenografts, but are resistant to trametinib (Fig. 1d). To investigate the cause of this slow-growth phenotype, we performed western blotting for a number of cell-cycle-associated proteins on parental A375 cells and EGFR-expressing derivatives. EGFR expression resulted in hypophosphorylated pRB protein, induction of the CDK inhibitors CDKN1A (p21^{cip1}) and CDKN1B (p27^{kip1}) and acidic β -galactosidase (Fig. 1e, f), markers that have been associated with oncogene-induced senescence^{15,16}. These markers were also induced upon expression of oncogenic versions of BRAF or MEK, but much less when activated mutants of AKT1 or PIK3CA were expressed in A375 cells (Extended Data Fig. 1). We conclude that EGFR expression is disadvantageous for BRAF(V600E) melanoma cells in the absence of BRAF or MEK inhibitor drugs, but it confers a selective advantage in the presence of these drugs.

Acquired EGFR expression may be the result of an adaptive response of the cancer cell population during drug selection. To ask in an unbiased way which factors might modulate EGFR expression in melanoma cells, we compiled a 'chromatin regulator' library of shRNAs targeting 661 genes, including the lysine acetyltransferases (KATs), lysine methyltransferases (KMTs), lysine deacetylases (KDACs), lysine demethylases (KDMs), chromatin remodelling complexes and proteins that harbour chromatin binding/associated domains (Supplementary Table 2). A375 melanoma cells, which express very low levels of EGFR, were infected with the chromatin regulator library and selected with vemurafenib for 3 weeks. Then the vemurafenib-resistant cells were collected and strongly EGFR-positive cells (EGFR^{hi}) were isolated from the drug-resistant

¹Division of Molecular Carcinogenesis, Cancer Systems Biology Centre and Cancer Genomics Centre Netherlands, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands. ²Department of Biochemistry, The Rosalind and Morris Goodman Cancer Centre, McGill University, Montreal, Quebec H3G 1Y6, Canada. ³Institut Gustave Roussy, 114 Rue Edouard Vaillant, 94800 Villejuif, France. ⁴Division of Medical Oncology, Cancer Systems Biology Centre and Cancer Genomics Centre Netherlands, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands. ⁵Division of Pathology, Cancer Systems Biology Centre and Cancer Genomics Centre Netherlands, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands. ⁶Department of Pathology, University Medical Centre Utrecht, Heidelberglaan 100, 3584 CX, Utrecht, The Netherlands. ⁷University of Torino, Department of Oncology, Strada 142 Km 3.95, 10060 Candiolo, Torino, Italy. ⁸Candiolo Cancer Institute – FPO, IRCCS, Strada 142 Km 3.95, 10060 Candiolo, Torino, Italy. ⁹FIRC Institute of Molecular Oncology (IFOM), 20139 Milano, Italy.

*These authors contributed equally to this work.

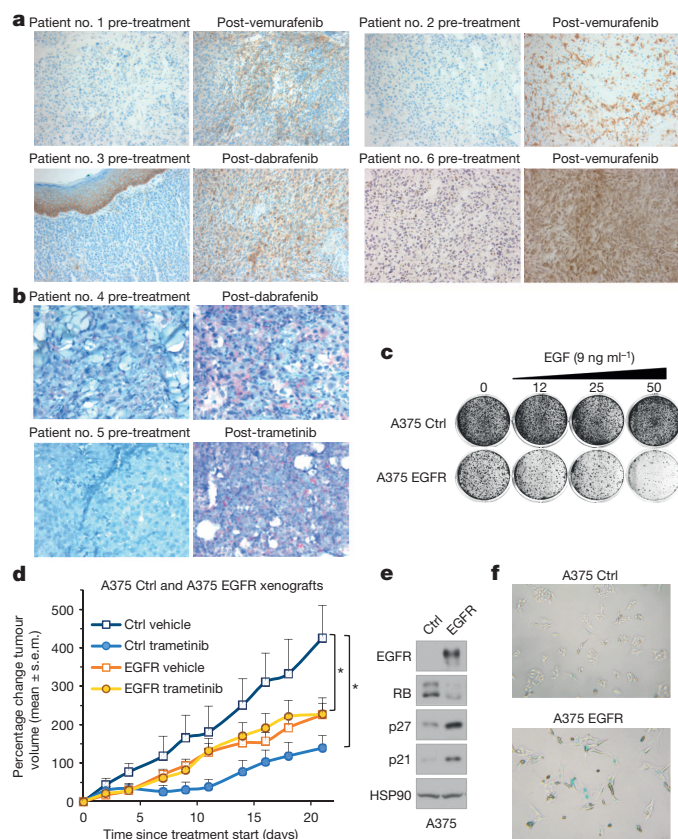


Figure 1 | Acquired *EGFR* expression in *BRAF(V600E)* mutant melanoma after vemurafenib resistance. **a, b**, Immunohistochemical (IHC) analysis (**a**, ultraView DAB stain, brown; **b**, ultraView Red stain, red) showing increased *EGFR* expression in formalin-fixed paraffin embedded (FFPE) (Patient number 1–5) and frozen (Patient number 6) melanoma tissue sections from *BRAF(V600E)* mutant melanoma patients who developed resistance to vemurafenib, dabrafenib or trametinib as indicated. For each patient, the first biopsy is from the pre-treatment tumour; the second biopsy was performed after the tumour had progressed under treatment. For patient number 4, the first biopsy was performed when the patient was in partial response, but rapidly developed secondary resistance. Then 4.5 months later, the second biopsy was taken. **c**, *EGFR* expression confers a growth disadvantage to *BRAF(V600E)* mutant melanoma cells and *EGFR* ligand potentiates the growth deficiency *in vitro*. A375 *BRAF(V600E)* melanoma cells transduced with control lentiviral vectors (Ctrl, pLX304-GFP) or vectors expressing *EGFR* (*EGFR*, pLX304-*EGFR*) were seeded at the same density and cultured in the presence of EGF at indicated concentration for 2 weeks. The cells were fixed, stained and photographed. **d**, *EGFR* expression confers a growth disadvantage to *BRAF(V600E)* mutant melanoma, but induces trametinib resistance *in vivo*. CD1 nude mice were inoculated with *BRAF(V600E)* mutant melanoma A375 cells transduced with control retroviral vectors or vectors expressing *EGFR*. Once tumours were established, animals were treated with vehicle or trametinib. Relative tumour volume is shown. Error bars represent s.e.m. ($n = 5$). $*P < 0.05$, single-sided Wilcoxon–Mann–Whitney test. **e**, Western blot analysis of RB protein, CDK inhibitors CDKN1A ($p21^{cip1}$) and CDKN1B ($p27^{kip1}$) in *EGFR* expressing A375 cells. HSP90 served as a loading control. **f**, *EGFR* expression induces senescence. Senescence was detected by staining of β -galactosidase activity. All experiments shown, except the ones that involve clinical samples and animals, were performed independently at least 3 times.

population by fluorescence-activated cell sorting (FACS) using an anti-*EGFR* antibody (Fig. 2a). Treatment of cells with either the chromatin regulator library or vemurafenib alone did not increase the fraction of *EGFR*^{hi} cells. In contrast, a significant fraction of *EGFR*^{hi} cells could be retrieved when cells were infected with the chromatin regulator library and were selected for vemurafenib resistance (Fig. 2b). We conclude that *EGFR*^{hi} melanoma cells do not merely appear as a consequence of silencing of certain chromatin regulators, but that these cells only emerge

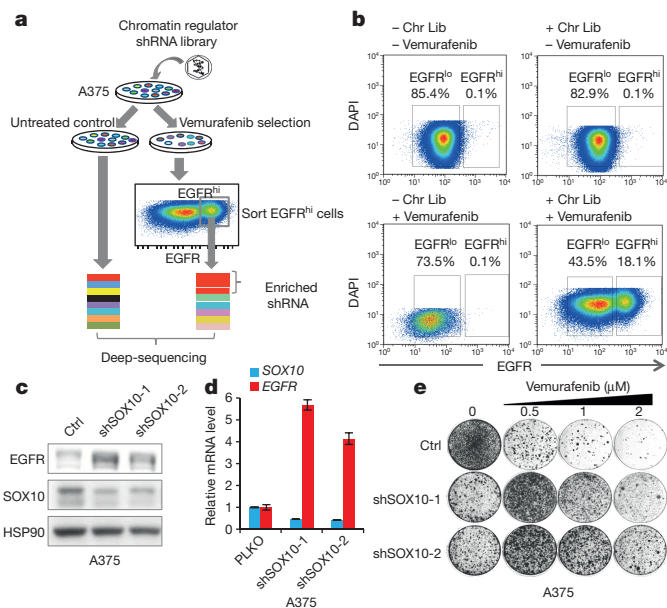


Figure 2 | FACS-assisted shRNA genetic screen identifies *SOX10* as a determinant of vemurafenib resistance and *EGFR* expression. **a**, Schematic outline of the FACS-assisted shRNA screen. Human ‘chromatin regulator’ shRNA library polyclonal virus was generated to infect A375 cells, which were then left untreated (control) or treated with 0.5 μ M vemurafenib. After 12 days, the untreated cells were collected. The cells that survived from 21 days of vemurafenib treatment were FACS sorted for *EGFR* expression. Subsequently, shRNA inserts from both samples were recovered by PCR and identified by massive parallel sequencing. **b**, *EGFR*^{hi} cells result from the combination of infection with chromatin regulator library and vemurafenib selection. A375 cells infected with ‘chromatin regulator’ library (Chr Lib) were cultured in the presence of 0.5 μ M vemurafenib for 21 days (right lower panel). Cells were harvested with 2 mM EDTA, stained with anti-*EGFR* antibody and analysed for *EGFR*^{hi} cells by flow cytometry. A375 cells cultured with or without vemurafenib, and A375 cells infected by the ‘chromatin regulator’ library without vemurafenib treatment served as controls. **c, d**, Suppression of *SOX10* induces *EGFR* expression. **c**, Western blot analysis of *EGFR* and *SOX10* levels in cells targeted by two independent sh*SOX10* vectors, HSP90 served as a loading control. **d**, The level of *EGFR* induction was determined by quantitative PCR with reverse transcription (qRT-PCR) analysis of the relative mRNA level of *EGFR*. pLKO.1 empty vector served as a control vector (Ctrl). Error bars represent s.d. of measurement replicates ($n = 3$). **e**, Two independent shRNAs targeting *SOX10* confer a proliferation disadvantage in the absence of drug, but induce vemurafenib resistance. A375 cells expressing shRNAs (as shown in Fig. 2c) targeting *SOX10* were seeded at the same density in 6-well plates and cultured in the absence (for 2 weeks) or presence of vemurafenib (for 4 weeks) at the indicated concentrations. The cells were fixed, stained and photographed. All experiments shown except shRNA screen were performed independently at least 3 times.

when the population is placed under drug-selection pressure. This indicates that silencing of the gene(s) that induce *EGFR* expression is not favoured in the absence of vemurafenib.

To identify which gene(s) in the chromatin regulator library when suppressed can induce *EGFR* expression, we isolated genomic DNA from the *EGFR*^{hi} cells and non-drug-treated control cells and determined the abundance of the shRNA vectors in each cell population by deep sequencing, as described previously³. shRNAs that confer resistance to vemurafenib through upregulation of *EGFR* should be enriched in the *EGFR*^{hi} fraction. shRNA screens are notorious for yielding false positive results. Therefore, in principle only those genes that are represented by multiple shRNAs should be followed up in a genetic screen¹⁷. However, in this screen we did not identify any genes for which more than one shRNA was enriched (Supplementary Table 3). We therefore focused on the top 10 most strongly enriched genes for follow-up experiments. We tested several additional shRNA vectors for each of

these 10 genes for their ability to increase *EGFR* expression, as this was a selection criterion in the genetic screen (Extended Data Fig. 2a, b). Only suppression of the SRY (sex determining region Y)-box 10 (*SOX10*) gene induced prominent *EGFR* expression when multiple *SOX10* shRNAs (*shSOX10*) were used in four melanoma cell line models (Fig. 2c, d and Extended Data Figs 2c, 4c and 5c). *SOX10* knockdown (*SOX10^{KD}*) induced a slow-growth phenotype and also displayed the hallmarks of oncogene-induced senescence in three melanoma models (Fig. 2e and Extended Data Figs 2e, f, g, 4b, e, f and 5b, e, f).

Next we confirmed that *SOX10^{KD}* indeed induced vemurafenib resistance in melanoma. We infected A375 cells with *shSOX10* and cultured cells in the presence of vemurafenib. *SOX10^{KD}* slowed down proliferation of A375 cells in the absence of drug, but in the presence of vemurafenib *SOX10^{KD}* conferred drug resistance in both short-term and long-term assays (Fig. 2e and Extended Data Fig. 2d, e). Moreover, under vemurafenib selective pressure, cells having a higher degree of *SOX10^{KD}* were selected, which consequently also expressed higher levels of *EGFR*, consistent with the notion that increased *EGFR* levels drive drug resistance (Extended Data Fig. 2h). Vemurafenib resistance through *SOX10* suppression was also seen in additional melanoma cell lines (Extended Data Figs 4a and 5a). A low concentration of vemurafenib actually increased the proliferation rate of *SOX10^{KD}* cells, consistent with the model that hyperactive BRAF–MEK signalling induces senescence markers, a process which is inhibited by vemurafenib (Extended Data Fig. 4a, g).

To study how *SOX10* suppression induces *EGFR* expression, we performed transcriptome sequencing (RNA-seq) of both parental A375 and A375 *SOX10^{KD}* cells (Supplementary Table 4). Gene set enrichment analysis of the *SOX10*-upregulated genes revealed an enrichment of genes with SMAD2/3 (downstream mediators of TGF- β signalling) and JUN binding sites in their promoters (Supplementary Table 5). Consistent with this, *SOX10* suppression induced TGF- β receptor 2 (*TGFBR2*) expression as well as a number of bona fide TGF- β target genes, including *JUN*, in three melanoma cell models (Fig. 3a, b and Extended Data Figs 4d and 5d). Levels of active JUN (phosphorylated JUN, p-JUN) were also increased by *SOX10^{KD}* (Fig. 3a). That treatment of melanoma cells with recombinant TGF- β causes resistance to vemurafenib further supports a role for TGF- β signalling in vemurafenib resistance (Fig. 3c and ref. 18). TGF- β 1 treatment caused not only induction of *EGFR* expression, but also induction of *PDGFRB* (Fig. 3d, e) and also resulted in induction of senescence-associated β -galactosidase (Fig. 3f). Consistently, *SOX10* suppression also induced *PDGFRB* expression (Extended Data Figs 3c, 4c and 5c). Moreover, suppression of *TGFBR2* inhibited *EGFR* and *PDGFRB* induction in *SOX10^{KD}* cells (Figs 3g, h), whereas ectopic expression of *TGFBR2* induced p-JUN, *EGFR* and *PDGFRB* expression (Fig. 3i). JUN is a regulator of *EGFR* expression and TGF- β regulates *PDGFRB*^{22,23}. *SOX10* is known to regulate the melanocyte transcription factor MITF²⁴. Indeed, A375 cells with *shSOX10* also had reduced MITF expression, but MITF suppression alone did not change *EGFR* or *PDGFRB* expression and did not cause vemurafenib resistance (Extended Data Fig. 7c–e). Our data provide support for a model in which activation of TGF- β signalling by *SOX10* loss leads to increased *EGFR* and *PDGFRB* expression and vemurafenib resistance.

Treatment of A375-*SOX10^{KD}* cells with a combination of both vemurafenib and the EGFR inhibitor gefitinib did not lead to proliferation arrest, indicating that EGFR was not the sole driver of drug resistance in *SOX10^{KD}* cells (Extended Data Fig. 3a). Indeed, an unbiased survey of RTKs revealed that *SOX10^{KD}* activated not only EGFR, but also PDGFRB and ERBB3 (Extended Data Fig. 3b, c). A similar pattern of RTK activation was observed following TGF- β 1 treatment, highlighting the similarity between *SOX10* suppression and acquired TGF- β signalling (Extended Data Fig. 3b, d). Many RTKs share two major downstream signalling pathways (RAS–MEK–ERK and PI3K–AKT). Consistent with this, we found that combined inhibition of these two downstream pathways using BRAF and PI3K inhibitors

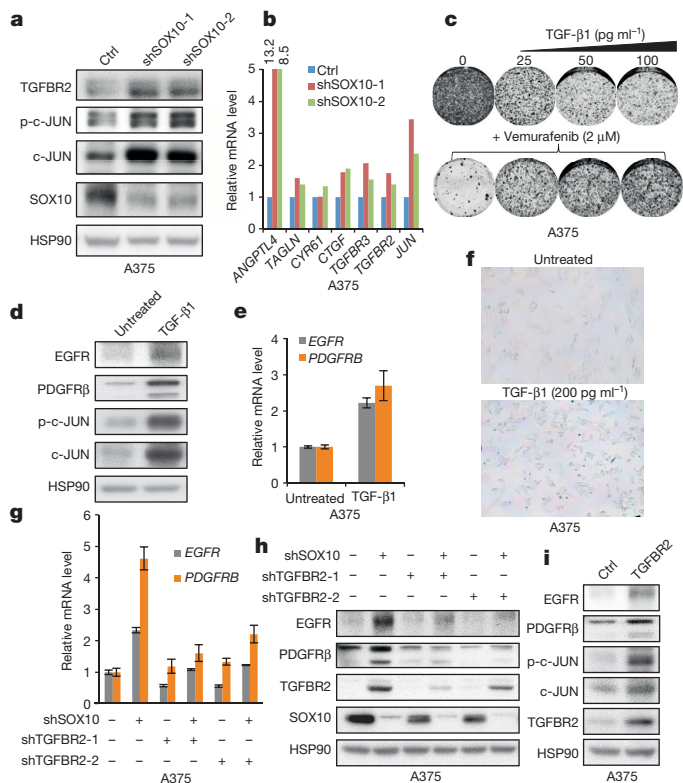


Figure 3 | Activation of TGF- β signalling leads to increased *EGFR* and *PDGFRB* expression. **a**, Suppression of *SOX10* activates TGFBR and JUN signalling. Two independent shRNAs targeting *SOX10* were individually introduced into A375 cells by lentiviral transduction. The levels of TGFBR2, p-JUN and JUN were determined by western blot analysis. HSP90 served as a loading control. **b**, *SOX10* loss leads to upregulation of TGF- β receptors and its bona fide target genes. Relative mRNA level of *ANGPTL4*, *TAGLN*, *CYR61*, *CTGF*, *TGFBR3*, *TGFBR2* and *JUN* were determined by transcriptome sequencing. pLKO.1 empty vector served as a control vector (Ctrl). **c**, TGF- β activation confers a growth disadvantage but vemurafenib resistance. A375 cells were seeded at the same density in 6-well plates and cultured in the absence or presence of recombinant TGF- β or vemurafenib at the indicated concentrations. The cells were fixed, stained and photographed. **d**, **e**, Recombinant TGF- β 1 treatment activates JUN and upregulates EGFR and PDGFRB expression. A375 cells were cultured in the absence or presence of 200 pg ml⁻¹ recombinant TGF- β 1 for 7 days before harvested for western blot or qRT-PCR analysis. Error bars represent s.d. of measurement replicates ($n = 3$). **f**, Recombinant TGF- β 1 treatment induces senescence. A375 cells were cultured in the presence of 200 pg ml⁻¹ recombinant TGF- β for 14 days. Senescence was detected by staining of β -galactosidase activity. **g**, **h**, *SOX10* loss induced EGFR and PDGFRB upregulation is TGFBR2-dependent. A375 cells were infected with lentiviral shRNA vectors as indicated. Relative mRNA levels of EGFR and PDGFRB were determined by qRT-PCR analysis; EGFR, PDGFRB, TGFBR2 and SOX10 levels were determined by western blot analysis. Error bars represent s.d. of replicate measurements ($n = 3$). **i**, TGFBR2 overexpression is sufficient to upregulate EGFR and PDGFRB. TGFBR2 was introduced to A375 cells by lentiviral transduction (TGFBR2, pLX304-TGFBR2). pLX304-GFP served as a control vector (Ctrl). The levels of EGFR, PDGFRB, TGFBR2, p-JUN and JUN were determined by western blot analysis. All experiments shown except RNA-seq were performed independently at least 3 times.

could restore growth inhibition in *SOX10^{KD}* cells (Extended Data Fig. 3a).

Our data are consistent with a model in which cells with low *SOX10* and high *EGFR* and *PDGFRB* expression are positively selected in the presence of drug, but that such cells are counter-selected in the absence of drug. To test this model directly, we infected A375 cells with *shSOX10* and subjected this heterogeneous population of *SOX10^{KD}* cells to vemurafenib selection for one week. At this point, we collected part of this population and determined EGFR expression by FACS analysis. Under

vemurafenib selection, an increased level of EGFR and a markedly decreased level of *SOX10* were observed. When these cells were subsequently cultured for one more week in the absence of vemurafenib, the EGFR^{hi}SOX10^{lo} population was depleted (Fig. 4a and Extended Data Fig. 6a). These data indicate that acquired EGFR expression is only advantageous to melanoma cells in the presence of drug selection, but is counter-selected in the absence of drug.

Consistent with a role for *SOX10* in regulation of EGFR expression in melanoma, we found an inverse correlation between *SOX10* and EGFR expression in a panel of 34 melanoma cell lines²⁵ (Fig. 4b) and a similar inverse relationship between *SOX10* and PDGFRB (Extended Data Fig. 6b). The most extreme cell line in this panel, LOXIMVI, completely lacked *SOX10* expression and had the highest EGFR expression. When we expressed *SOX10* in this cell line, EGFR and PDGFRB were reduced and TGFBR2 and TGFBR3 as well as JUN and p-JUN levels were also downregulated, consistent with the notion that *SOX10* regulates these RTKs through an effect on TGF- β signalling (Extended

Data Fig. 6c, d). Consistently, expression of *SOX10* in LOXIMVI cells increased their sensitivity to vemurafenib (Extended Data Fig. 6e).

To ask directly whether *SOX10* is involved in EGFR-associated drug resistance in *BRAF*(V600E) melanoma patients, we isolated RNA from the six patients studied above who had gained EGFR expression after acquisition of trametinib, dabrafenib or vemurafenib resistance (Supplementary Table 1). We performed RNA-seq analysis to determine changes in transcriptome upon drug resistance. In two patients, the levels of *SOX10* mRNA were reduced (Fig. 4c and Extended Data Fig. 6f). EGFR and PDGFRB mRNA were greatly increased in patient 5, whereas no evidence was found in this patient of alternative *BRAF* splicing⁷ or *BRAF* overexpression (Extended Data Fig. 7a, b). Patient 3 had strong induction of EGFR protein post-resistance (Fig. 1a), but at first glance, EGFR mRNA levels appeared only minimally induced. However, scrutiny of the RNA-seq data revealed that the apparent lack of induction of EGFR in this tumour sample pair is caused by the abnormally high EGFR transcript abundance in the pre-treatment

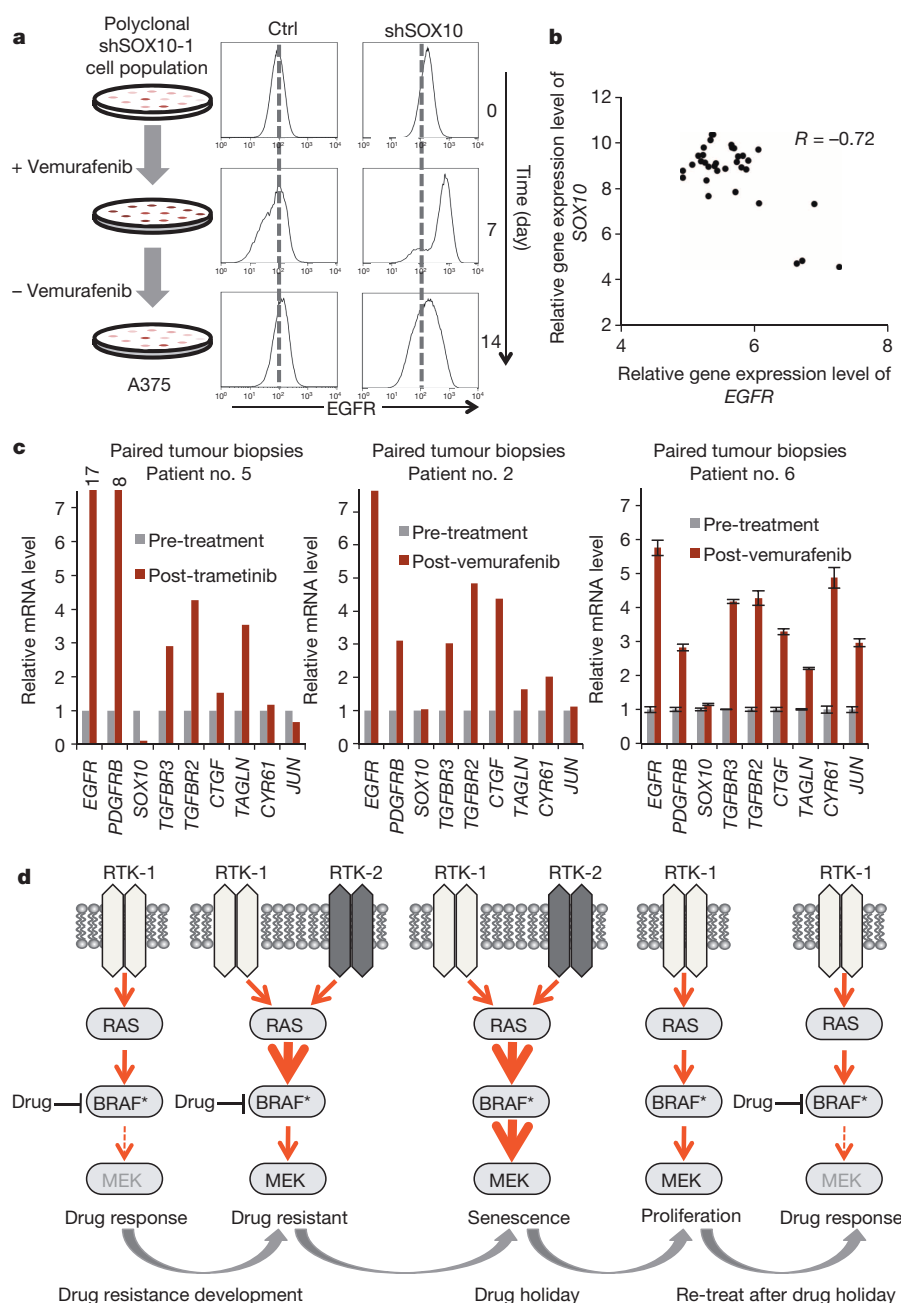


Figure 4 | Inverse relationship between *SOX10* and receptor tyrosine kinase expression in melanoma. **a**, Intermittent drug dosing alters relative proportions of EGFR^{hi} and EGFR^{lo} cell populations. A375 cells were infected with shSOX10-1 to generate a polyclonal cell population of SOX10^{KD} cells. The infected cells were seeded in 6-well plates, samples collected and stained with antibody against EGFR for flow cytometry analysis at day 0, day 7 and day 14 (0.5 μ M vemurafenib treatment started on day 0 and stopped on day 7). pLKO.1 (Ctrl) vector served as a control. **b**, Inverse correlation between *SOX10* and EGFR in a panel of human *BRAF* mutant melanoma cell lines. Relative gene expression levels of *SOX10* and EGFR were acquired from the Cancer Cell Line Encyclopedia (CCLE). R stands for Pearson product-moment correlation coefficient. **c**, Differential gene expression of *SOX10*, EGFR, PDGFRB, TGF- β receptors and TGF- β target genes in pre- and post-treatment patient tumour biopsies. Total RNA was isolated from FFPE specimens derived from tumour biopsies of patient 5, 2 and 6 both before and after development of drug resistance. After reverse transcription, gene expression levels were determined by transcriptome sequencing (patient 5 and patient 2) qRT-PCR analysis (patient 6). Error bars represent s.d. of measurement replicates ($n = 3$). **d**, Model for senescence induction after development of vemurafenib resistance. Upregulation of RTKs leads to enhanced signalling through the RAS-BRAF-MEK pathway. Consequently, vemurafenib is no longer able to fully silence the signalling to MEK and drug resistance is seen. When the drug is removed, supra-physiological levels of BRAF-MEK signalling induced a state of oncogene-induced senescence, which subsequently leads to negative selection of the RTKs and restores drug responsiveness. All experiments shown, except the ones that involve clinical samples, were performed independently at least 3 times.

sample and not the lack of *EGFR* expression in the post-treatment sample (Extended Data Fig. 6g). This is most probably owing to the contamination of this sample with the strongly *EGFR*-positive skin material (see Fig. 1a). These tumours also manifested increased TGF- β signalling (Fig. 4c and Extended Data Fig. 6h). Two further pairs of tumour samples showed induction of *EGFR* and *PDGFRB* without notable loss of *SOX10* after drug resistance emerged. These tumours displayed induction of TGF- β receptor expression and induction of a number of bona fide TGF- β targets, indicating that these tumours somehow had acquired TGF- β signalling (and subsequent induction of *EGFR* and *PDGFRB* expression) in a *SOX10*-independent fashion (Fig. 4c).

Clinical evidence indicates that melanoma patients that have developed vemurafenib resistance can regain sensitivity to the drug after a drug holiday, suggesting a reversible and adaptive transcriptional response to the drug²⁶. That drug resistance is reversed in the absence of drug indicates that this adaptive response is not favoured in the absence of drug. Our data provide a molecular underpinning for the concept that drug resistance may arise at a fitness cost in the absence of drug (Fig. 4d). Melanoma patients whose tumours acquire *EGFR* expression as a result of drug resistance development may be candidates to be re-treated with drug after a drug holiday.

METHODS SUMMARY

Pooled shRNA Screen. A 'chromatin regulator' shRNA library targeting 661 genes was constructed from the TRC human genome-wide shRNA collection (TRC-Hs1.0). Lentiviral shRNA vectors generated from the pooled library were used to infect A375 cells. Cells stably expressing shRNA were selected by vemurafenib and then FACS sorted for *EGFR* expression. Massive parallel sequencing was used to determine the enriched shRNA in the selected cell population.

Melanoma patient tumour samples. Permission was granted by the NKI and IGR ethical committee to take biopsies from *BRAF*(V600E) mutant patients before and after vemurafenib, dabrafenib or trametinib treatment. All patients consented to participate in the study. *BRAF*(V600E) mutation status was determined by Departments of Pathology at NKI and IGR.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 22 April 2013; accepted 31 January 2014.

Published online 26 March 2014.

- Chapman, P. B. *et al.* Improved survival with vemurafenib in melanoma with *BRAF* V600E mutation. *N. Engl. J. Med.* **364**, 2507–2516 (2011).
- Flaherty, K. T. *et al.* Improved survival with MEK inhibition in *BRAF*-mutated melanoma. *N. Engl. J. Med.* **367**, 107–114 (2012).
- Prahalad, A. *et al.* Unresponsiveness of colon cancer to *BRAF*(V600E) inhibition through feedback activation of *EGFR*. *Nature* **483**, 100–103 (2012).
- Corcoran, R. B. *et al.* *EGFR*-mediated re-activation of MAPK signaling contributes to insensitivity of *BRAF* mutant colorectal cancers to RAF inhibition with vemurafenib. *Cancer Discov.* **2**, 227–235 (2012).
- Davies, H. *et al.* Mutations of the *BRAF* gene in human cancer. *Nature* **417**, 949–954 (2002).
- Flaherty, K. T., Hodi, F. S. & Fisher, D. E. From genes to drugs: targeted strategies for melanoma. *Nature Rev. Cancer* **12**, 349–361 (2012).
- Poulikakos, P. I. *et al.* RAF inhibitor resistance is mediated by dimerization of aberrantly spliced *BRAF*(V600E). *Nature* **480**, 387–390 (2011).

- Johannessen, C. M. *et al.* COT drives resistance to RAF inhibition through MAP kinase pathway reactivation. *Nature* **468**, 968–972 (2010).
- Wagle, N. *et al.* Dissecting therapeutic resistance to RAF inhibition in melanoma by tumour genomic profiling. *J. Clin. Oncol.* **29**, 3085–3096 (2011).
- Shi, H. *et al.* Melanoma whole-exome sequencing identifies (V600E)*BRAF* amplification-mediated acquired *BRAF* inhibitor resistance. *Nature Commun.* **3**, 724 (2012).
- Nazarian, R. *et al.* Melanomas acquire resistance to *BRAF*(V600E) inhibition by RTK or *N-RAS* upregulation. *Nature* **468**, 973–977 (2010).
- Girotti, M. R. *et al.* Inhibiting *EGF* receptor or *SRC* family kinase signaling overcomes *BRAF* inhibitor resistance in melanoma. *Cancer Discov.* **3**, 158–167 (2013).
- Villanueva, J. *et al.* Acquired resistance to *BRAF* inhibitors mediated by a RAF kinase switch in melanoma can be overcome by cotargeting MEK and IGF-1R/PI3K. *Cancer Cell* **18**, 683–695 (2010).
- Real, F. X. *et al.* Expression of epidermal growth factor receptor in human cultured cells and tissues: relationship to cell lineage and stage of differentiation. *Cancer Res.* **46**, 4726–4731 (1986).
- Serrano, M., Lin, A. W., McCurrach, M. E., Beach, D. & Lowe, S. W. Oncogenic *ras* provokes premature cell senescence associated with accumulation of p53 and p16^{INK4a}. *Cell* **88**, 593–602 (1997).
- Michaloglou, C. *et al.* *BRAF*^{E600}-associated senescence-like cell cycle arrest of human naevi. *Nature* **436**, 720–724 (2005).
- Brummelkamp, T. R. & Bernards, R. New tools for functional mammalian cancer genetics. *Nature Rev. Cancer* **3**, 781–789 (2003).
- Huang, S. *et al.* MED12 controls the response to multiple cancer drugs through regulation of TGF- β receptor signaling. *Cell* **151**, 937–950 (2012).
- Johnson, A. C. *et al.* Activator protein-1 mediates induced but not basal epidermal growth factor receptor gene expression. *Mol. Med.* **6**, 17–27 (2000).
- Zenz, R. *et al.* c-Jun regulates eyelid closure and skin tumor development through *EGFR* signaling. *Dev. Cell* **4**, 879–889 (2003).
- Steller, E. J. *et al.* *PDGFRB* promotes liver metastasis formation of mesenchymal-like colorectal tumor cells. *Neoplasia* **15**, 204–217 (2013).
- Zhang, Y., Feng, X.-H. & Derynck, R. Smad3 and Smad4 cooperate with c-Jun/c-Fos to mediate TGF- β -induced transcription. *Nature* **394**, 909–913 (1998).
- Mialon, A. *et al.* DNA topoisomerase I is a cofactor for c-Jun in the regulation of epidermal growth factor receptor expression and cancer cell proliferation. *Mol. Cell. Biol.* **25**, 5040–5051 (2005).
- Bondurand, N. *et al.* Interaction among *SOX10*, *PAX3* and *MITF*, three genes altered in Waardenburg syndrome. *Hum. Mol. Genet.* **9**, 1907–1917 (2000).
- Garnett, M. J. *et al.* Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* **483**, 570–575 (2012).
- Seghers, A. C., Wilgenhof, S., Lebbe, C. & Neyns, B. Successful rechallenge in two patients with *BRAF*-V600-mutant melanoma who experienced previous progression during treatment with a selective *BRAF* inhibitor. *Melanoma Res.* **22**, 466–472 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the NKI Core Facilities for Genomics and Molecular Pathology & Biobanking for tumour tissue and support in DNA sequencing. We thank S. Roy for collecting clinical data and N. Kamsu Kom for tissue preparation. This work was supported by grants from the European Research Council (ERC), the Dutch Cancer Society (KWF), the EU COLTHERES project and grants by the Netherlands Organization for Scientific Research (NWO) to Cancer Genomics Netherlands (CGC.NL). Additional support was provided by Fondazione Piemontese per la Ricerca sul Cancro—ONLUS grant 'Farmacogenomica—5 per mille 2009 MIUR' (F.D.N.); AIRC MFAG 11349 (F.D.N.); AIRC IG grant n. 12812 (A.B.); and Canadian Institutes of Health Research (CIHR) grant MOP-130540 (S.Hu).

Author Contributions R.B., A.B., F.D.N., L.F.A.W., C.R., R.L.B. and A.M.M.E. supervised all research. R.B. and C.S. wrote the manuscript. C.S., L.W., S.Hu., G.J.J.E.H., A.P., D.Z., S.Ho., P.K.B., C.L., C.M., S.V., J.W., W.G., I.H. and A.S. designed and performed experiments and J.H., C.B., C.R., S.M.W., S.V. and A.M.M.E. provided clinical samples and gave advice.

Author Information RNA sequencing data are available at Gene Expression Omnibus with accession code GSE50535. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.B. (r.bernards@nki.nl).

Maternal retinoids control type 3 innate lymphoid cells and set the offspring immunity

Serge A. van de Pavert^{1†*}, Manuela Ferreira^{2*}, Rita G. Domingues², Hélder Ribeiro², Rosalie Molenaar¹, Lara Moreira-Santos², Francisca F. Almeida², Sales Ibizá², Inês Barbosa², Gera Goverse¹, Carlos Labão-Almeida², Cristina Godinho-Silva², Tanja Konijn¹, Dennis Schooneman¹, Tom O'Toole¹, Mark R. Mizee¹, Yasmin Habani¹, Esther Haak³, Fabio R. Santori⁴, Dan R. Littman⁴, Stefan Schulte-Merker⁵, Elaine Dzierzak³, J. Pedro Simas², Reina E. Mebius^{1*} & Henrique Veiga-Fernandes^{2*}

The impact of nutritional status during fetal life on the overall health of adults has been recognized¹; however, dietary effects on the developing immune system are largely unknown. Development of secondary lymphoid organs occurs during embryogenesis and is considered to be developmentally programmed^{2,3}. Secondary lymphoid organ formation depends on a subset of type 3 innate lymphoid cells (ILC3) named lymphoid tissue inducer (LTi) cells²⁻⁵. Here we show that mouse fetal ILC3s are controlled by cell-autonomous retinoic acid (RA) signalling *in utero*, which pre-sets the immune fitness in adulthood. We found that embryonic lymphoid organs contain ILC progenitors that differentiate locally into mature LTi cells. Local LTi cell differentiation was controlled by maternal retinoid intake and fetal RA signalling acting in a haematopoietic cell-autonomous manner. RA controlled LTi cell maturation upstream of the transcription factor RORγt. Accordingly, enforced expression of *Rorgt* restored maturation of LTi cells with impaired RA signalling, whereas RA receptors directly regulated the *Rorgt* locus. Finally, we established that maternal levels of dietary retinoids control the size of secondary lymphoid organs and the efficiency of immune responses in the adult offspring. Our results reveal a molecular link between maternal nutrients and the formation of immune structures required for resistance to infection in the offspring.

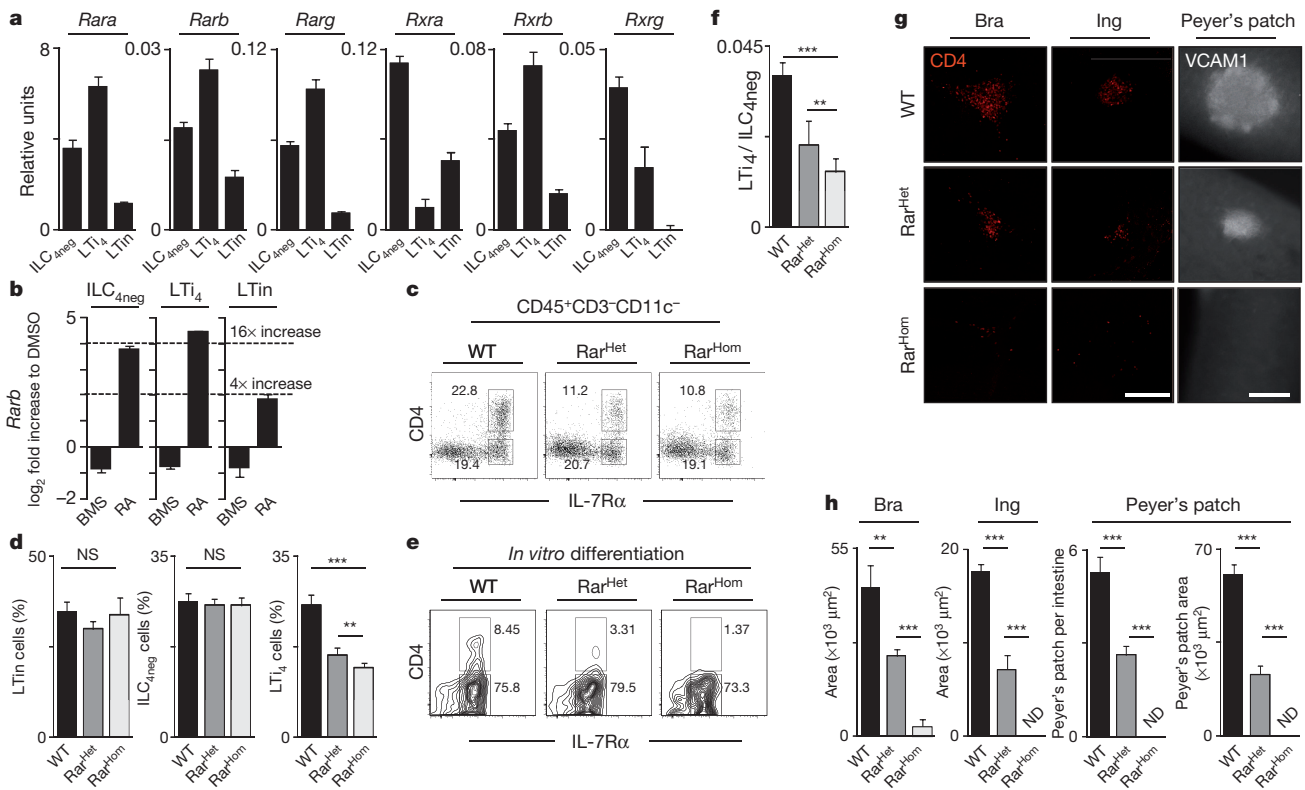
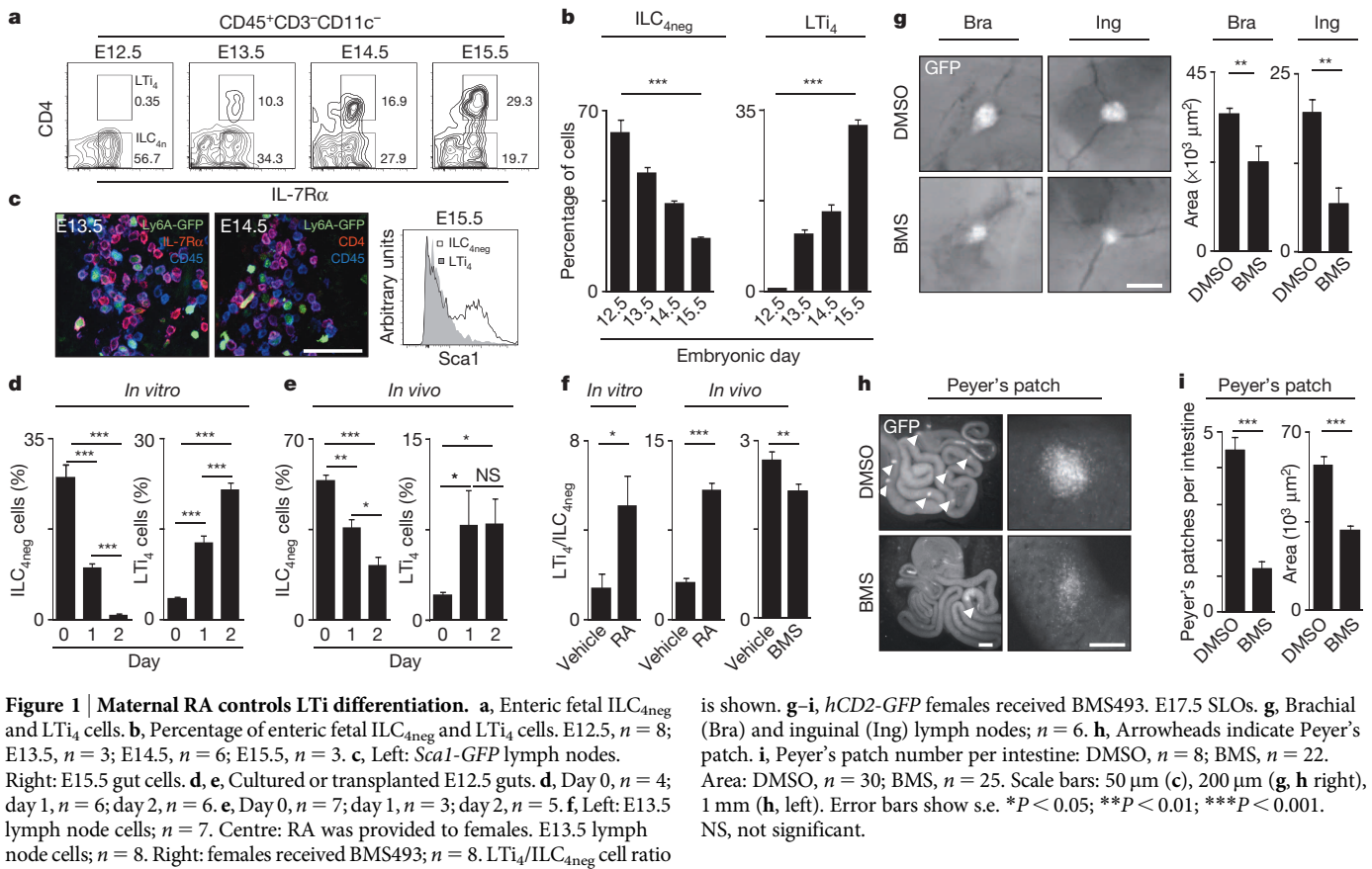
Haematopoietic cells that initially colonize secondary lymphoid organ (SLO) sites include CD3[−]c-Kit⁺IL-7Rα[−]α4β7⁺CD11c[−]CD4[−] lymphoid tissue initiator (LTin) cells and the prototypical member of type 3 ILCs, LTi cells²⁻⁷. Although most LTi cells express CD4, this is a late event in LTi differentiation and not all RORγt⁺ LTi cells express this marker^{5,6,8,9}. Thus, we proposed that CD3[−]IL-7Rα[−]α4β7⁺ID2⁺c-Kit⁺CD11c[−]CD4[−] ILCs (hereafter called ILC_{4neg} cells) receive local cues giving rise to ID2⁺RORγt⁺CD4⁺ LTi cells (LTi₄) within developing SLOs. Notably, enteric ILC_{4neg} cells include mainly ID2⁺RORγt⁺CD4[−] LTi cells (LTi₀) but also a small fraction of ID2⁺RORγt[−]CD4[−] precursors with LTi cell potential (hereafter called pre-ILC cells)⁹. In contrast, nearly 100% of lymph node ILC_{4neg} cells are LTi₀ cells (Extended Data Fig. 1a, b). Analysis of embryonic day 12.5 (E12.5) guts revealed that ILC_{4neg} cells are the only appreciable IL-7Rα⁺ colonizing cells (Fig. 1a, b). Accordingly, non-cycling mature Sca1[−]LTi₄ cells increased throughout development, seemingly at the expense of Sca1⁺ILC_{4neg} cells (Fig. 1a–c and Extended Data Fig. 1c). Further evidence that ILC_{4neg} cells differentiate locally was provided by organ cultures and transplantation of E12.5 intestines. Despite absence of fetal liver output in these settings, LTi₄ cells increased with time at the expense of local ILC_{4neg} cells (Fig. 1d, e). Furthermore, in E14.5 *Rorgt*^{−/−} embryos, ILC_{4neg} cells were attracted to the intestine and lymph nodes, supporting initial anlagen colonization by these cells (Extended Data Fig. 1d, e).

Notably, RA stimulation of E13.5 lymph node cells showed increased frequency of LTi₄ cells and reduction of ILC_{4neg} cells, indicating that differentiation of LTi₄ cells is regulated by RA (Fig. 1f). To confirm the effect of RA in LTi differentiation *in vivo*, pregnant mice received a RA-enriched diet starting at E10.5. Supplementation of RA increased the proportion of LTi₄ cells in the embryo, to the detriment of ILC_{4neg} cells (Fig. 1f). In agreement with this finding, provision of the RA signalling inhibitor BMS493 to pregnant female mice resulted in a decrease of fetal LTi₄ cells despite normal frequency of fetal liver progenitors and enteric haematopoietic cells (Fig. 1f and Extended Data Fig. 1f). Consequently, despite normal embryo size, BMS493 administration led to a reduction in lymph node dimensions and Peyer's patch developmental failure (Fig. 1g–i and Extended Data Fig. 1g). Collectively, our data indicate that maternal retinoids control LTi cell differentiation within developing SLOs.

RA is a vitamin A metabolite that controls early vertebrate development, some immune processes in adulthood, and has been shown to mediate CXCL13 expression in fetal mesenchymal cells^{10–16}. RA binds to heterodimers formed by the RA receptors (RARs) and retinoid X receptors (RXRs), which bind DNA RA response elements (RAREs)¹¹. To address putative RA cell-autonomous responses, we assessed RAR and RXR expression in E15.5 ILC_{4neg}, LTi₄ and LTin cells. RARs and RXRs were predominantly expressed by ILC_{4neg} and LTi₄ cells, whereas LTin cells expressed these molecules at lower levels (Fig. 2a). RA stimulation revealed that only ILC_{4neg} and LTi₄ cells respond robustly, as shown by *Rarb* upregulation (Fig. 2b)¹⁶. Together, these data indicate that impaired SLO development in BMS493-treated mice might be the consequence of RA signal ablation in LTi cells. To test this hypothesis we used a lineage-targeted model to block RA signalling. We used a mouse line in which a truncated form of the RARα gene was knocked into the ROSA26 locus preceded by a triple polyadenylation signal flanked by two loxP sites (ROSA26-RARα403). This line was bred to Vav-iCre mice that in contrast to other tested Cre lines ensured Cre activity in fetal LTin, ILC_{4neg} and LTi₄ cells (Extended Data Fig. 2a–d)^{17,18}. Despite normal frequencies of fetal liver precursors, and SLO LTin, ILC_{4neg} and LTi₀ cells, Vav-iCre/ROSA26-RARα403 embryos (Rar mice) revealed a dose-dependent reduction of LTi₄ cells (Fig. 2c, d and Extended Data Fig. 3a–f). To assess whether the differentiation potential of ILC_{4neg} cells is controlled by RA thresholds, we cultured purified ILC_{4neg} cells from Rar heterozygous (Rar^{Het}), homozygous (Rar^{Hom}) and wild-type littermate control mice. Whereas wild-type ILC_{4neg} cells upregulated pro-inflammatory cytokines and chemokines and gave rise to LTi₄ cells *in vitro*, these were impaired proportionally to the degree of RA signalling abrogation in ILC_{4neg} cells (Fig. 2e, f and Extended Data Fig. 3g). Finally, despite normal frequency of colonizing ILC_{4neg} cells (Fig. 2d),

¹Department of Molecular Cell Biology and Immunology, VU University Medical Center, van der Boechorststraat 7, 1081BT Amsterdam, The Netherlands. ²Instituto de Medicina Molecular, Faculdade de Medicina de Lisboa, Av. Prof. Egas Moniz, Edifício Egas Moniz, 1649-028 Lisboa, Portugal. ³Erasmus Stem Cell Institute, Department of Cell Biology, Erasmus Medical Center, 3000 CA Rotterdam, The Netherlands. ⁴Howard Hughes Medical Institute, Molecular Pathogenesis Program, Skirball Institute of Biomolecular Medicine, New York University School of Medicine, New York, New York 10016, USA. ⁵Hubrecht Institute–KNAW (Royal Netherlands Academy of Arts and Sciences) and University Medical Center Utrecht, 3584 CT Utrecht, Netherlands. †Present address: Hubrecht Institute–KNAW (Royal Netherlands Academy of Arts and Sciences) and University Medical Center Utrecht, 3584 CT Utrecht, Netherlands.

*These authors contributed equally to this work.



haematopoietic cell-autonomous impairment of RA responses resulted in severely diminished fetal lymph node size and reduced number of minute Peyer's patches (Fig. 2g, h and Extended Data Fig. 3h). Our data indicate that LT_i cell differentiation is controlled by cell-autonomous RA signalling in developing SLOs.

Previous reports have identified key LT_i cell regulators². Mice mutant for the transcription factors ID2 and ROR γ t lack LT_i cells and do not develop SLOs^{19,20}. *Runx1*, *Tox* and *Notch1* were also implicated in LT_i cell maturation^{9,21–23}. We found that whereas most LT_i-related genes were normally expressed in Rar^{Hom} and Rar^{Het} ILC_{4neg} and LT_{i4} cells, *Runx1* was increased and *Rorgt* was reduced (Fig. 3a and Extended Data Fig. 4a–d). Expression of pro-inflammatory genes was also reduced in Rar^{Hom} and Rar^{Het} ILC_{4neg} and LT_{i4} cells (Fig. 3a and Extended Data Fig. 4b–d). The marked reduction of *Rorgt* expression suggested that RA could provide ILC_{4neg} cells with signals leading to *Rorgt* regulation. Accordingly, RA stimulation of ILC_{4neg} cells resulted in *Rorgt* upregulation whereas most other transcription factors were unperturbed, notably *Runx1* (Fig. 3b). In agreement, BMS493 inhibited RA-induced *Rorgt* expression, and efficient block of ROR γ t by digoxin prevented RA-induced differentiation of ILC_{4neg} cells into LT_{i4} cells, while cell viability was unaffected (Fig. 3c and Extended Data Fig. 5a–c). To test further whether RA-induced LT_i maturation requires ROR γ t, we determined if differentiation of RAR dominant-negative ILC_{4neg} cells is restored by enforced *Rorgt* expression. Retroviral transduction of *Rorgt* revealed that RAR dominant-negative ILC_{4neg} cells restored high levels of pro-inflammatory genes and reacquired their potential to differentiate towards LT_{i4} cells (Fig. 3d–f). Further evidence that RA can directly regulate *Rorgt* expression was provided by computational analysis of potential RARE sites and chromatin immunoprecipitation (ChIP) with pan-RAR and RXR antibodies. RA stimulation resulted in increased binding of RAR and RXR upstream and within the *Rorgt* locus (Fig. 3g, h and Extended Data Table 1). To analyse the role of these sites we introduced the RARE C (–5,478 *Rorgt* transcription start site (TSS)), E (–1,800 *Rorgt* TSS) and G (–1,619 *Rorgt* TSS) half-sites in a luciferase reporter vector. Mutations in these sites resulted in significant reduction of the regulatory function of these elements as measured by luciferase activity (Fig. 3i). Thus, cell-autonomous RA signalling provides LT_i cells with critical differentiation signals via direct regulation of *Rorgt*.

Our data indicate that mature LT_i cell numbers regulate the size of SLO primordia and may determine lymphoid organ size in adulthood²⁴. Rar^{Het} adult mice had reduced SLOs and lymphocyte numbers when compared to their wild-type littermate controls (Fig. 4a, b and Extended Data Fig. 6a, b). In agreement, mice that received a vitamin-A-deficient (VAD) diet throughout life had reduced lymphoid organ size when compared to vitamin-A-control mice (VAC) (Fig. 4c). However, because Rar^{Het} and VAD lymphocytes are continuously exposed to altered levels of RA signals, it is possible that SLO size might be a consequence of altered lymphocyte pools^{12–15}. To clarify this issue, we provided pregnant mice with a vitamin-A-high (VAH), VAD or VAC diet and switched all diets to the same VAC diet after birth. At 10 weeks of age mice that were exposed to a VAH diet exclusively *in utero* had larger SLOs, whereas mice exposed to a VAD diet had small SLOs when compared to VAC control mice (Fig. 4d). Notably, provision of variable vitamin A diet levels exclusively after birth no longer controlled SLO size (Extended Data Fig. 6c, d). Additional evidence that RA determines SLO size in early life was provided by transplantation of CD45.1 wild-type bone marrow into lethally irradiated Rar^{Het} (WT→Rar^{Het}) or wild-type (WT→WT) CD45.2 littermate control hosts at 2 weeks of age (Fig. 4e). Thus, we generated mice that pre- and perinatally received low input of RA signals in haematopoietic cells, but that on transplantation harbour a normal wild-type haematopoietic system. WT→Rar^{Het} mice, which received low RA cues *in utero*, exhibited small SLOs when compared to their WT→WT counterparts at 8 weeks after transplantation (Fig. 4f, g and Extended Data Fig. 6e, f). This phenotype also revealed reduced lymphocyte numbers, albeit normal SLO organization and similar haematopoietic cell reconstitution (Fig. 4h and Extended Data Figs 6f and 7a–d). In agreement, dendritic cells from WT→Rar^{Het} or WT→WT chimaeras had similar capacity to activate lymphocytes (Extended Data Fig. 7e, f).

Our data indicate that available RA *in utero* regulates the size of lymphocyte pools in the offspring, with possible consequences on their adaptive immune responses. To test this hypothesis, WT→Rar^{Het} or WT→WT chimaeras were infected intranasally with murine herpesvirus-4, resulting in acute lung infection. Analysis of draining intrathoracic lymph nodes revealed reduced expansion but normal frequency of CD8⁺ T cells specific for the viral epitopes ORF61 and ORF75c in WT→Rar^{Het} mice (Fig. 4i and Extended Data Fig. 8a–c). Consequently, whereas

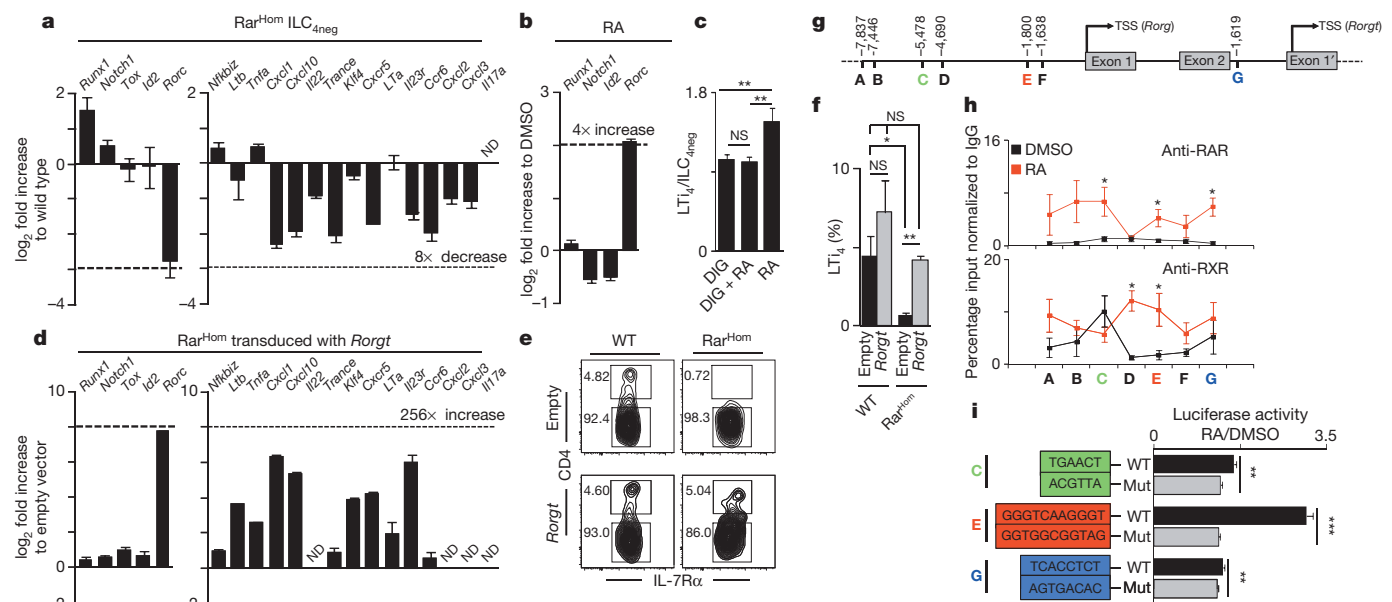


Figure 3 | RA controls LT_i cells via ROR γ t. **a**, E15.5 ILC_{4neg} cells. Data represent three experiments. **b**, E15.5 ILC_{4neg} cells stimulated with RA. Data represent three experiments. **c**, E13.5 lymph node cells stimulated with RA and digoxin (DIG); $n = 16$. **d–f**, ILC_{4neg} cells transduced with pMig. *Rorgt*-IRES-GFP virus (day 6). **d**, Data represent three experiments.

e, Cytometry analysis. **f**, Emergent LT_{i4} cells. Wild type (WT), $n = 5$; Rar^{Hom}, $n = 3$. **g**, RARE sites. **h**, E15.5 ILC_{4neg} cells stimulated with RA. ChIP analysis of five biological replicates. **i**, Luciferase activity of RARE wild-type and mutated sites. C, $n = 6$; E, $n = 9$; G, $n = 6$. Error bars show s.e. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$. NS, not significant. ND, not detected.

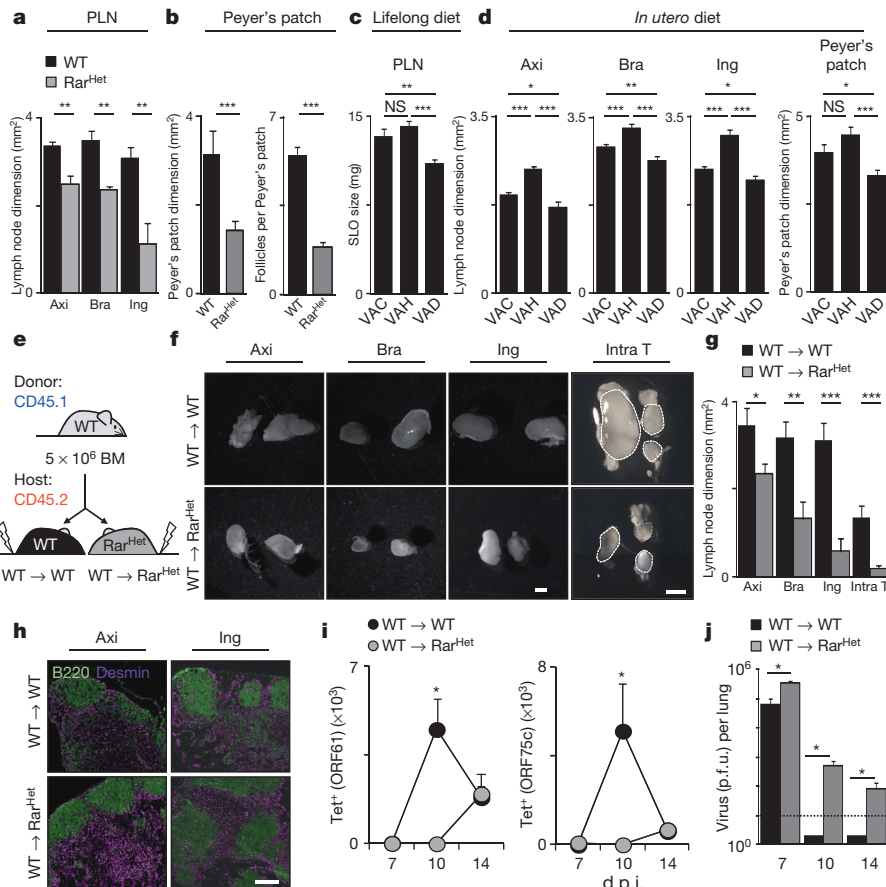


Figure 4 | Retinoid levels *in utero* determine the offspring resistance to infection. **a**, Adult axillary (Axi), brachial (Bra) and inguinal (Ing) peripheral lymph nodes (PLN). $n = 6$. **b**, Peyer's patch area and follicle number per Peyer's patch. Wild type (WT), $n = 26$; Rar^{Het}, $n = 23$. **c**, Females received diets that were maintained in the offspring. VAC, $n = 10$; VAH, $n = 8$; VAD, $n = 18$. **d**, Females received variable diets. Their offspring received VAC diet. Lymph node: VAC, $n = 18$; VAH, $n = 24$; VAD, $n = 25$. Peyer's patch: VAC, $n = 20$; VAH, $n = 40$; VAD, $n = 63$. **e**, Transplantation scheme. **f**, Chimaeric lymph

nodes. Scale bar: 1 mm. **g**, Dimensions of chimaeric lymph nodes. $n = 6$. **h**, Chimaeric lymph node structure. Scale bar: 200 μ m. **i**, Chimaeras infected with murine herpesvirus-4. Tetramer positive CD8 T cells in intrathoracic lymph nodes. **j**, Virus titres (plaque-forming units (p.f.u.) per lung). 7 days post infection (d.p.i.): WT → WT, $n = 3$; WT → Rar^{Het}, $n = 3$. 10 d.p.i.: WT → WT, $n = 4$; WT → Rar^{Het}, $n = 3$. 14 d.p.i.: WT → WT, $n = 8$; WT → Rar^{Het}, $n = 5$. Dashed line indicates the detection limit. Error bars show s.e. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$. NS, not significant.

WT → WT chimaeras efficiently cleared lytic virus by day 10, high viral titres were still detected in WT → Rar^{Het} chimaeras 14 days after infection (Fig. 4j)²⁵. Thus, a deficit of RA signals within haematopoietic cells in early life results in small SLOs and poor capacity to control infection.

Defining the requirements that control SLO development is essential to understand how immunity may be regulated. We now show that RA controls LT_i cells, regulating LT_i pro-inflammatory genes and the frequency of LT₄ cells (Extended Data Fig. 9). RA operates in a cell-autonomous fashion, via ROR γ t, although additional factors regulate Ror γ t in ILC3s.

SLO development has been considered to be developmentally programmed; we now show that formation of these structures can be also controlled by dietary signals. Thus, in addition to the established impact of dietary plant-derived chemicals in postnatal immune cells, our work reveals dietary retinoids as key regulators of pre-natal ILCs with a lifelong impact on adult lymphoid organ size^{26–28}. It was previously shown that complete absence of lymphoid organs leads to long-life virus persistence²⁹. Our data reveal that the efficiency of adaptive immune responses to infection and possibly to other immune insults may be pre-tuned in early life through dietary signals from maternal origin.

We report here that cell-autonomous RA signalling is a key axis for Ror γ t expression and LT_i cell differentiation within developing SLOs. Similarly, RA may also be important after birth in infection and chronic inflammatory diseases³⁰. Lineage-targeted strategies will be central to elucidate the contribution of dietary retinoids in these outcomes.

METHODS SUMMARY

Mice were maintained at Instituto de Medicina Molecular (IMM) or VU University Medical Centre according to national and international guidelines. Bone marrow cells were isolated from 8-week-old C57BL/6 CD45.1 mice and injected intravenously into 2-week-old lethally irradiated CD45.2 ROSA26-RAR α 403^{Het} (WT → Rar^{Het}) or wild-type littermate controls (WT → WT). C57BL/6 female mice received either vitamin-A-deficient (VAD), with no vitamin A (vitamin free casein), vitamin-A-high (VAH, 25,000 IU kg⁻¹) or vitamin-A-control (VAC, 4,000 IU kg⁻¹) diets. Retinoic acid was provided to pregnant mice from E10.5 until they were euthanized. Quantitative real-time PCR with reverse transcription (RT-PCR) was performed as previously described^{6,7,10}. Computational analysis was performed with TESS. DNA-protein complexes were immunoprecipitated using antibodies against mouse pan-RAR, pan-RXR or control IgG.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 6 June 2013; accepted 18 February 2014.

Published online 19 March 2014.

- Gluckman, P. D. & Hanson, M. A. Living with the past: evolution, development, and patterns of disease. *Science* **305**, 1733–1736 (2004).
- van de Pavert, S. A. & Mebius, R. E. New insights into the development of lymphoid tissues. *Nature Rev. Immunol.* **10**, 664–674 (2010).
- Randall, T. D., Carragher, D. M. & Rangel-Moreno, J. Development of secondary lymphoid organs. *Annu. Rev. Immunol.* **26**, 627–650 (2008).
- Mebius, R. E., Rennert, P. & Weissman, I. L. Developing lymph nodes collect CD4⁺CD3⁺LT β ⁺ cells that can differentiate to APC, NK cells, and follicular cells but not T or B cells. *Immunity* **7**, 493–504 (1997).

5. Eberl, G. *et al.* An essential function for the nuclear receptor ROR γ t in the generation of fetal lymphoid tissue inducer cells. *Nature Immunol.* **5**, 64–73 (2004).
6. Veiga-Fernandes, H. *et al.* Tyrosine kinase receptor RET is a key regulator of Peyer's Patch organogenesis. *Nature* **446**, 547–551 (2007).
7. Patel, A. *et al.* Differential RET signaling pathways drive development of the enteric lymphoid and nervous systems. *Sci. Signal.* **5**, ra55 (2012).
8. Cupedo, T. *et al.* Presumptive lymph node organizers are differentially represented in developing mesenteric and peripheral nodes. *J. Immunol.* **173**, 2968–2975 (2004).
9. Cherrier, M., Sawa, S. & Eberl, G. Notch, Id2, and ROR γ t sequentially orchestrate the fetal development of lymphoid tissue inducer cells. *J. Exp. Med.* **209**, 729–740 (2012).
10. van de Pavert, S. A. *et al.* Chemokine CXCL13 is essential for lymph node initiation and is induced by retinoic acid and neuronal stimulation. *Nature Immunol.* **10**, 1193–1199 (2009).
11. Niederreither, K. & Dolle, P. Retinoic acid in development: towards an integrated view. *Nature Rev. Genet.* **9**, 541–553 (2008).
12. Iwata, M. Retinoic acid production by intestinal dendritic cells and its role in T-cell trafficking. *Semin. Immunol.* **21**, 8–13 (2009).
13. Hall, J. A. *et al.* Essential role for retinoic acid in the promotion of CD4⁺ T cell effector responses via retinoic acid receptor alpha. *Immunity* **34**, 435–447 (2011).
14. Mora, J. R. & von Andrian, U. H. Role of retinoic acid in the imprinting of gut-homing IgA-secreting cells. *Semin. Immunol.* **21**, 28–35 (2009).
15. Mucida, D. *et al.* Retinoic acid can directly promote TGF- β -mediated Foxp3⁺ Treg cell conversion of naive T cells. *Immunity* **30**, 471–472; Reply 472–473 (2009).
16. Hall, J. A., Grainger, J. R., Spencer, S. P. & Belkaid, Y. The role of retinoic acid in tolerance and immunity. *Immunity* **35**, 13–22 (2011).
17. de Boer, J. *et al.* Transgenic mice with hematopoietic and lymphoid specific expression of Cre. *Eur. J. Immunol.* **33**, 314–325 (2003).
18. Rosselot, C. *et al.* Non-cell-autonomous retinoid signaling is crucial for renal development. *Development* **137**, 283–292 (2010).
19. Sun, Z. *et al.* Requirement for ROR γ in thymocyte survival and lymphoid organ development. *Science* **288**, 2369–2373 (2000).
20. Yokota, Y. *et al.* Development of peripheral lymphoid organs and natural killer cells depends on the helix-loop-helix inhibitor Id2. *Nature* **397**, 702–706 (1999).
21. Aliahmad, P., de la Torre, B. & Kaye, J. Shared dependence on the DNA-binding factor TOX for the development of lymphoid tissue-inducer cell and NK cell lineages. *Nature Immunol.* **11**, 945–952 (2010).
22. Possot, C. *et al.* Notch signaling is necessary for adult, but not fetal, development of ROR γ t⁺ innate lymphoid cells. *Nature Immunol.* **12**, 949–958 (2011).
23. Tachibana, M. *et al.* Runx1/Cbfb2 complexes are required for lymphoid tissue inducer cell differentiation at two developmental stages. *J. Immunol.* **186**, 1450–1457 (2011).
24. Meier, D. *et al.* Ectopic lymphoid-organ development occurs through interleukin 7-mediated enhanced survival of lymphoid-tissue-inducer cells. *Immunity* **26**, 643–654 (2007).
25. Gredmark-Russ, S., Cheung, E. J., Isaacson, M. K., Ploegh, H. L. & Grotenbreg, G. M. The CD8 T-cell response against murine gammaherpesvirus 68 is directed toward a broad repertoire of epitopes from both early and late antigens. *J. Virol.* **82**, 12205–12212 (2008).
26. Kiss, E. A. *et al.* Natural aryl hydrocarbon receptor ligands control organogenesis of intestinal lymphoid follicles. *Science* **334**, 1561–1565 (2011).
27. Lee, J. S. *et al.* AHR drives the development of gut ILC22 cells and postnatal lymphoid tissues via pathways dependent on and independent of Notch. *Nature Immunol.* **13**, 144–151 (2011).
28. Qiu, J. *et al.* The aryl hydrocarbon receptor regulates gut immunity through modulation of innate lymphoid cells. *Immunity* **36**, 92–104 (2011).
29. Karrer, U. *et al.* On the key role of secondary lymphoid organs in antiviral immune responses studied in alymphoplastic (*aly/aly*) and spleenless (*Hox11^{-/-}*) mutant mice. *J. Exp. Med.* **185**, 2157–2170 (1997).
30. Spencer, S. P. *et al.* Adaptation of innate lymphoid cells to a micronutrient deficiency promotes type 2 barrier immunity. *Science* **343**, 432–437 (2014).

Acknowledgements We thank the imaging, animal and flow cytometry facilities at IMM and UPC for technical assistance; C. Mendelsohn for providing ROSA26-RAR α 403 mice; N. Schmölke, J. G. van Rietschoten, R. E. van Kesteren, T. H. B. Geijtenbeek, S. Gringhuis, E. Keuning, J. Peterson-Maduro, M. G. Roukens, D. D'Astolfo, M. Vermunt, A. Rijkerkerk, J. Koning, J. van der Meulen and B. Oliver for technical help; and G. Vilhais-Neto, M. C. Coles and G. Eberl for discussion. M.F., L.M.-S. and R.G.D. were supported by FCT, Portugal; H.V.-F. by EMBO (1648) and ERC (207057); D.R.L. by NIH (RO1AI080885) and HHMI; M.R.M. by Dutch MS research foundation (MS 12-797); S.A.vd.P. by NGL Breakthrough Horizon (40-41009-98-9077); and R.E.M. by a VICI (918.56.612) and ALW-TOP grant (09.048).

Author Contributions M.F. wrote the manuscript, designed, performed and analysed the experiments in Figs 1a, b, d, e, g–i, 2a–h, 3a, b, d–i and 4a, b, d–j and Extended Data Figs 1a, d–f, 2a–d, 3a–h, 4a–d, 5a–c, 6, 7a–f, 8a–d, 9a, b and 10a–c. S.A.vd.P. wrote the manuscript, designed, performed and analysed the experiments in Figs 1c, f, 3c, g–i and 4c and Extended Data Figs 1b, c, e, 5b, c and 6. R.G.D., H.R., R.M., L.M.-S., F.F.A., S.J., I.B., G.G., C.L.-A., T.K., D.S., T.O'T., M.R.M., Y.H. and S.S.-M. contributed to several experiments. C.G.-S. and J.P.S. provided murine herpesvirus-4. D.R.L. and F.R.S. provided *Ror γ t^{-/-}* embryos. E.H. and E.D. provided Ly-6A (*Sca1*)-GFP mice. R.E.M. and H.V.-F. supervised the work, planned the experiments and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.V.-F. (jhfernandes@medicina.ulisboa.pt).

Mitoflash frequency in early adulthood predicts lifespan in *Caenorhabditis elegans*

En-Zhi Shen^{1,2*}, Chun-Qing Song^{1,2*}, Yuan Lin^{3*}, Wen-Hong Zhang², Pei-Fang Su⁴, Wen-Yuan Liu², Pan Zhang², Jiejia Xu³, Na Lin³, Cheng Zhan², Xianhua Wang³, Yu Shyr⁵, Heping Cheng³ & Meng-Qiu Dong²

It has been theorized for decades that mitochondria act as the biological clock of ageing¹, but the evidence is incomplete. Here we show a strong coupling between mitochondrial function and ageing by *in vivo* visualization of the mitochondrial flash (mitoflash), a frequency-coded optical readout reflecting free-radical production and energy metabolism at the single-mitochondrion level^{2,3}. Mitoflash activity in *Caenorhabditis elegans* pharyngeal muscles peaked on adult day 3 during active reproduction and on day 9 when animals started to die off. A plethora of genetic mutations and environmental factors inversely modified the lifespan and the day-3 mitoflash frequency. Even within an isogenic population, the day-3 mitoflash frequency was negatively correlated with the lifespan of individual animals. Furthermore, enhanced activity of the glyoxylate cycle contributed to the decreased day-3 mitoflash frequency and the longevity of *daf-2* mutant animals. These results demonstrate that the day-3 mitoflash frequency is a powerful predictor of *C. elegans* lifespan across genetic, environmental and stochastic factors. They also support the notion that the rate of ageing, although adjustable in later life, has been set to a considerable degree before reproduction ceases.

Ageing is defined as an increase in the probability of death with increasing chronological age⁴. In eukaryotic organisms, the mitochondrion organelle is pivotal to bioenergetics, free-radical metabolism and cell death⁵. Conceptualized by Harman in 1972, the mitochondrial theory of ageing¹ (MTA) states that mitochondria are the main biological clock that drives ageing. However, it has been difficult to prove or disprove MTA and related theories because experimental approaches have had limitations⁶.

We recently demonstrated that mitochondria in mammalian cells undergo quantal, stochastic bursts of superoxide production, which can be visualized as 'mitochondrial flashes' (mitoflashes) by the sensor protein circularly permuted yellow fluorescent protein (cpYFP) in the mitochondrial matrix². The mitoflash frequency is highly sensitive to oxidative stress and metabolic changes³; it can therefore be exploited as a readout of energy metabolism and free-radical production for testing MTA.

We expressed cpYFP with a mitochondrion-targeting sequence⁷ (mt-cpYFP) in *C. elegans*, a model system long used for ageing studies⁸. The highest expression of mt-cpYFP was detected in the pharynx, a strong muscular structure that grinds bacterial food (Extended Data Fig. 1). For ease of detection, we focused on the anterior pharynx (procorpus and metacarpus). Robust mitoflash activity occurred in the pharyngeal muscle of wild-type (WT) worms (Fig. 1a and Supplementary Videos 1 and 2), with attributes (Extended Data Fig. 2) comparable to those of mammalian mitoflashes².

Systematic characterization further showed that mitoflashes are a conserved physiological phenomenon. First, *C. elegans* mutants with defective respiratory chain complexes all had extremely low mitoflash activity (Fig. 1b), substantiating the previous finding that pharmacological inhibition of the respiratory chain diminishes mitoflash activity

in mammalian cells². Second, glucose supplementation in the culture plates stimulated mitoflash production, whereas the removal of food for 3–6 h diminished it (Fig. 1c). Third, oxidative stressors such as H₂O₂ or paraquat (PQ), a herbicide that generates intracellular reactive oxygen species (ROS), almost immediately increased the mitoflash frequency about 15-fold (Fig. 1d and Supplementary Video 3). A similar increase was elicited by photochemically induced ROS from the fluorescent protein KillerRed⁹ (Fig. 1d). Thus, when the electron-transfer chain is intact, the mitoflash frequency in *C. elegans* pharyngeal muscles reflects both the metabolic state and oxidative stress, as in mammalian cells^{2,3}.

To determine the relationship between mitoflashing and ageing, we first monitored mitoflash activity in *C. elegans* as a function of age. The WT mitoflash frequency peaked on adult days 2–3 and again on days 8.5–9.5 at about four events per optical section of pharynx over a 200-s window (Fig. 1e and Supplementary Videos 1 and 2). The first peak overlapped with the egg-laying period spanning days 1–5, and it can be suppressed by eliminating the germline (Extended Data Fig. 3). The second peak coincided with the time when WT animals started to die off (Fig. 2a inset, grey curve). Next we measured mitoflash activity in long-lived and short-lived mutants during their adult life. In *C. elegans*, *daf-2* encodes the insulin/insulin-like growth factor-1 receptor, and *age-1* encodes a phosphoinositide 3-kinase; these act to limit lifespan by inhibiting DAF-16, a FOXO transcription factor⁸. In the WT background, *daf-2(e1370)* mutation and *age-1* RNA-mediated interference (RNAi) extended the average lifespan from 21 days to 39 and 30 days, respectively (Fig. 2a, b insets). The first mitoflash peak in both mutants was depressed and delayed to day 5; the second peak, strikingly higher than the first, did not appear until day 15 or 13 (Fig. 2a, b). The longevity of the *daf-2* mutant is known to require the *daf-16* gene¹⁰. In the *daf-2; daf-16* double mutant, both the lifespan and the mitoflash time course were restored to a WT-like state (Fig. 2c). In contrast, mitoflashes in the short-lived *hsf-1(sy441)* and *elo-5(RNAi)* mutants, which have decreased activity of the heat-shock transcription factor or a polyunsaturated fatty acid elongase, showed an elevated first peak on day 3 and a compressed second peak around day 7 (Fig. 2d, e). Thus, from early to mid-adulthood the pace of mitoflashing seems to be slower in long-lived mutants and faster in short-lived mutants, suggesting a connection between mitoflashing and ageing.

Ageing is plastic: genetic, environmental and stochastic factors together determine how long an animal lives. We therefore examined whether mitoflash activity and ageing would display coherent responses to any of these factors, provided that the mitoflash was measured in healthy respiratory mitochondria, for example in day-3 young adults. We examined 29 genetic or RNAi mutant strains (excluding respiratory-chain mutants) with long, short or WT-like lifespans. This collection covered several evolutionarily conserved mechanisms that control lifespan⁸, including insulin signalling (*daf-2*, *age-1*, *daf-2; daf-16*), dietary restriction (DR) or TOR signalling (*eat-2*, *daf-15*, *pha-4*) and germline signalling

¹College of Biological Sciences, China Agricultural University, Beijing 100094, China. ²National Institute of Biological Sciences, Beijing, Beijing 102206, China. ³State Key Laboratory of Biomembrane and Membrane Biotechnology, Beijing Key Laboratory of Cardiometabolic Molecular Medicine, Institute of Molecular Medicine, Peking-Tsinghua Center for Life Sciences, Peking University, Beijing 100871, China. ⁴Department of Statistics, National Cheng Kung University, Tainan 70101, Taiwan. ⁵Vanderbilt Centre for Quantitative Sciences, Vanderbilt University, Nashville, Tennessee 37232, USA.

*These authors contributed equally to this work.

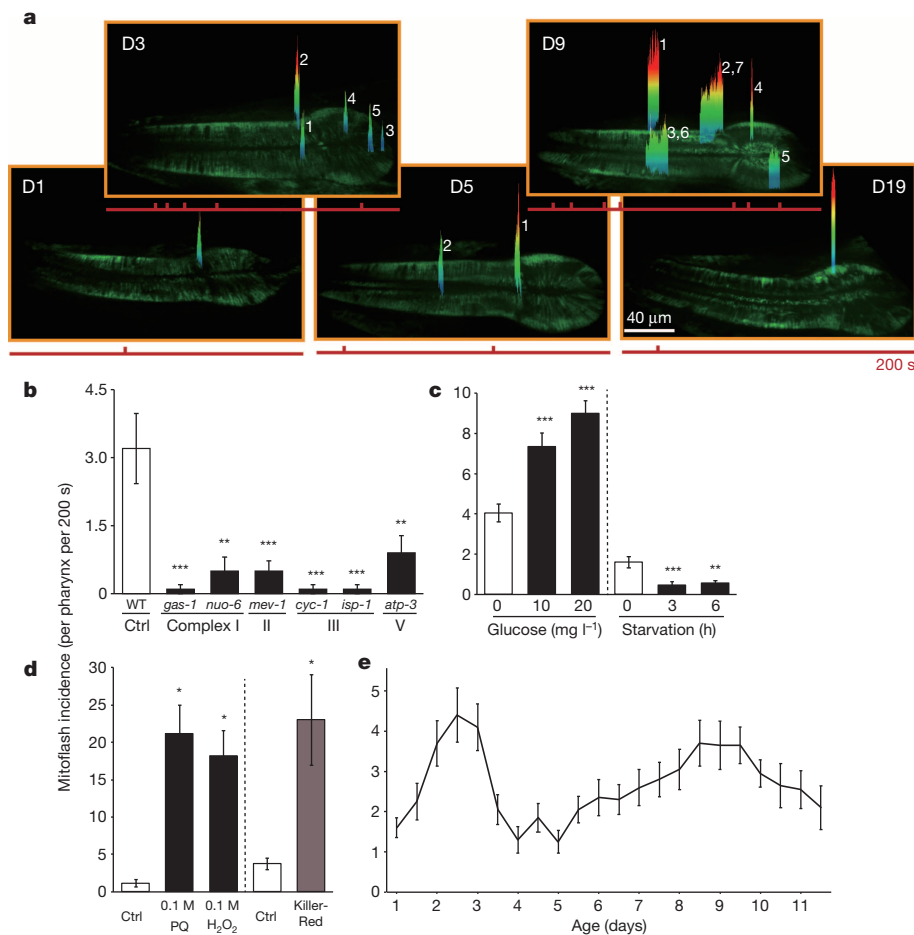


Figure 1 | Mitoflash events in mt-cpYFP-transgenic *C. elegans*. **a**, Space-time plots of mitoflash events in mt-cpYFP-transgenic *C. elegans* pharynx on adult days 1, 3, 5, 9 and 19 (D1, D3, D5, D9 and D19, respectively). The spatial localization of mitoflash events is shown by surface plots overlaying the respective confocal images of the pharynx. Numerals denote the order of occurrence when more than one mitoflash was registered in the optical section (1 μm thick). Temporal diaries of these events, marked as vertical ticks, during a 200-s acquisition window are shown beneath the images. **b**, Mitoflash frequencies in worms with defective respiratory chain components ($n = 10$). Mutants were *gas-1(fc21)*, *nuo-6(qm200)*, *mev-1(RNAi)*, *cyc-1(RNAi)*, *isp-1(qm150)* and *atp-3(RNAi)*. Ctrl, control. **c**, Mitoflash response to glucose supplementation or starvation ($n = 20$). **d**, Mitoflash response to treatment with paraquat (PQ) or H_2O_2 , or to light exposure in worms expressing mitochondrion-targeted KillerRed ($n = 5$). **e**, Age dependence of mitoflash activity in *C. elegans* ($n = 20$). In **b–d**, day-1 adults were used for the PQ, H_2O_2 and starvation experiments, and day-3 adults for the others. Data are expressed as means \pm s.e.m. Asterisk, $0.01 \leq P < 0.05$; two asterisks, $0.001 \leq P < 0.01$; three asterisks, $P < 0.001$; Wilcoxon rank-sum test.

(*glp-1*). When all the data were pooled, a strong negative correlation (Spearman's correlation: -0.82) emerged between the day-3 mitoflash frequency (varying from about 5% to 255% of the WT) and the mean lifespan (varying from about 57% to 191% of the WT) (Fig. 3a and Supplementary Tables 1 and 3). The variation in day-3 mitoflash frequency alone explained 59% of the variation in lifespan using a nonlinear

cubic spline regression model (Fig. 3b), or 50% using a simple linear regression model (Fig. 3c) ($P < 0.001$ for both). Hence the day-3 mitoflash frequency reflects the effect of genetic manipulation on ageing.

Next we sought to determine how the day-3 mitoflash frequency and the rate of ageing respond to diverse environmental factors. We tested 26 conditions that had been reported to alter the lifespan of *C. elegans*^{11–20}

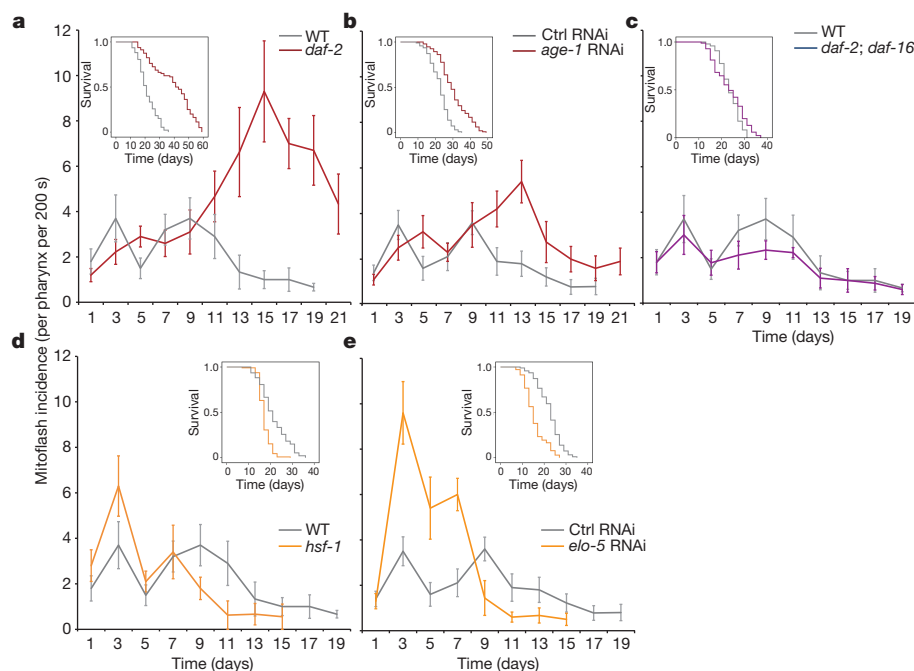
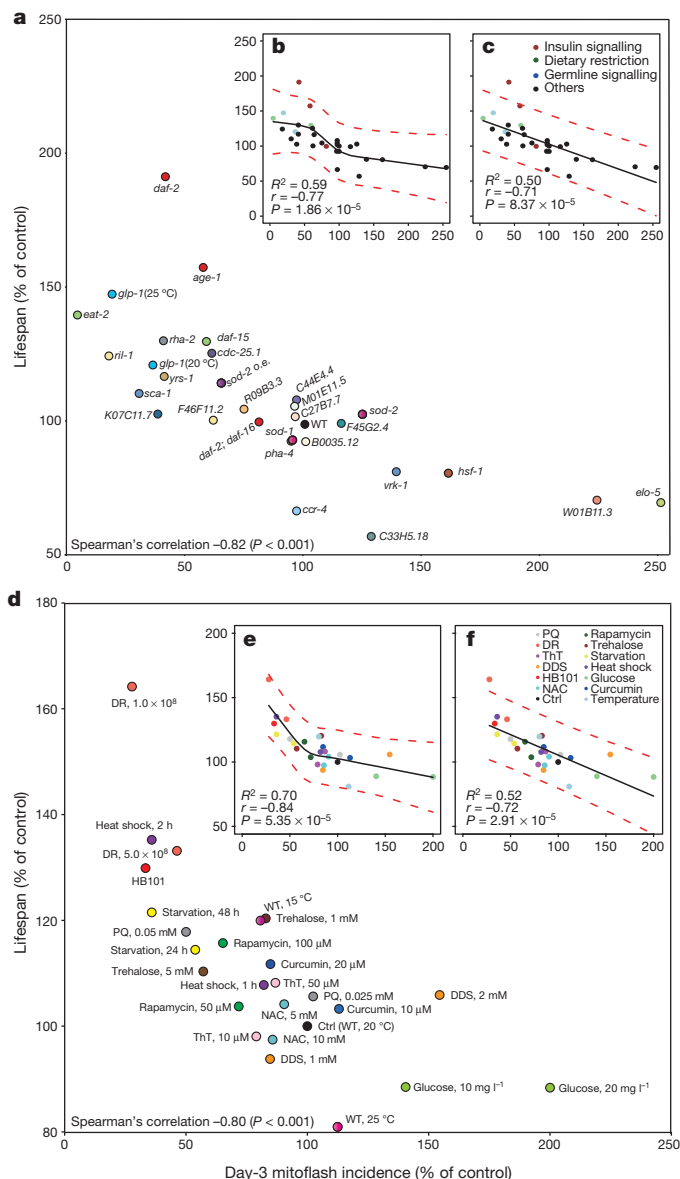


Figure 2 | Mitoflashes in lifespan mutants. Age-dependent changes in mitoflash activity in the long-lived insulin signalling mutants *daf-2(e1370)* (**a**) and *age-1(RNAi)* (**b**), the WT-like *daf-2(e1370); daf-16(mu86)* double mutant (**c**), and the short-lived mutants *hsf-1(sy441)* (**d**) and *elo-5(RNAi)* (**e**). Insets show the corresponding survival curves. For each mitoflash result (mean \pm s.e.m.), $n = 9$ or 10 worms except for the *daf-2* mutant on day 13, *age-1(RNAi)* on day 17, *daf-2; daf-16* on day 17, and *elo-5(RNAi)* on day 15, for which $n = 8$. For the lifespan curves in **a–e** respectively, $n = 80, 93, 85, 111$ and 103 worms for the mutants, or 93, 93, 98, 93 and 94 for the WT or WT on empty RNAi control.



(Fig. 3d and Supplementary Tables 2 and 3). Again, there was a strong negative correlation (Spearman's correlation: -0.80) between the day-3 mitoflash frequency (varying from 28% to 200% of the control) and mean lifespan (varying from 81% to 164% of the control). Most of the lifespan variation could be explained by the day-3 mitoflash frequency (52% using a linear regression model, or 70% using a cubic spline regression model; $P < 0.001$ for both) (Fig. 3e, f). Thus, the day-3 mitoflash frequency also reflects the environmental effect on ageing.

Of particular interest are transient environmental interventions. It has been shown that a temporary early-life experience, such as restricted milk intake imposed on a crowded litter of mice before weaning²¹ or a sublethal heat shock or starvation of *C. elegans* day-1 adults¹⁸, can significantly increase life expectancy. Hence, the rate of ageing can be modulated in young animals through gene–environment interactions.

Table 1 | Effects of transient treatments on long-lived or short-lived *C. elegans*

<i>daf-2(e1370)</i> animals, long-lived	100 mM PQ treatment				WT on 10 mg l ⁻¹ glucose, short-lived	Starvation	
	1 min	30 min	60 min	120 min		3 h	6 h
MF after treatment (on day 2)	473*	92	31**	13**	MF after starvation (on day 2)	56*	47**
MF after recovery (on day 3)	92	86	126	608**	MF after recovery (on day 3)	67	52#
Lifespan	101	87	65***	66***	Lifespan	111**	118***

Values are percentages of control. Hash sign, $0.05 \leq P < 0.1$; asterisk, $0.01 \leq P < 0.05$; two asterisks, $0.001 \leq P < 0.01$; three asterisks, $P < 0.001$; one-tailed Wilcoxon rank-sum test for mitoflash frequency (MF) ($n = 15$ for the *daf-2* MF data on day 3, $n = 10$ for all others), and log-rank for lifespan ($n \geq 44$; see Supplementary Table 3 for precise n).

Indeed, either 2 h of heat shock or 48 h of starvation starting from adult day 1 increased the lifespan and decreased the day-3 mitoflash frequency (Fig. 3d). Moreover, 1–2 h of PQ treatment on adult day 2 was sufficient to shorten the lifespan of the long-lived *daf-2(e1370)* mutant, accompanied by an elevated mitoflash frequency on adult day 3 (Table 1). Conversely, 3–6 h of starvation on day 2 significantly increased the lifespan of short-lived WT animals cultured on glucose plates and decreased the mitoflash frequency on day 3 (Table 1). These results show that the effects of previous transient environmental changes on day-3 mitoflash activity faithfully predict their lasting effects on lifespan.

Worms experiencing an environmental change can have different mitoflash frequencies during, immediately after, and long after the treatment (Table 1, PQ treatment of the *daf-2* mutant). The delayed responses (measured on day 3 after 22–24 h of recovery), rather than the acute ones, were inversely related to the lifespan alterations. Hence, for the day-3 mitoflash frequency to be a predictor of lifespan, the system must be given time to recover or reset after transient perturbations.

A population of isogenic *C. elegans* cultured under the same condition have different lifespans (for example, 11–36 days for WT animals in Fig. 2a) as a result of stochastic factors²². Given that the mitoflash frequency can predict lifespan changes by both genetic and environmental factors, it was even more intriguing to determine whether the mitoflash frequency could foretell the lifespan of individual animals, whose lifespan differences can only be attributed to random factors. A negative correlation between the day-3 mitoflash frequency and lifespan still held true in strains with a normal, faster or slower ageing rate (Extended Data Fig. 4a) ($R^2 = 0.28, 0.19$ and 0.29 for WT, long-lived *age-1(hx546)* and short-lived *elo-5(RNAi)* worms, respectively; $P < 0.001$, linear regression model). Tripartition of the data by mitoflash frequency revealed that worms with low mitoflash activity as a group lived significantly longer than those with medium or high mitoflash activity (Extended Data Fig. 4b).

In contrast, there was no correlation between lifespan and the day-9 mitoflash frequency in individual WT worms maintained under the same condition (Extended Data Fig. 5). This was probably due to the heterogeneity of day-9 mitochondria. Although mitoflash activity is indicative of metabolic rate or oxidative state in healthy respiratory mitochondria

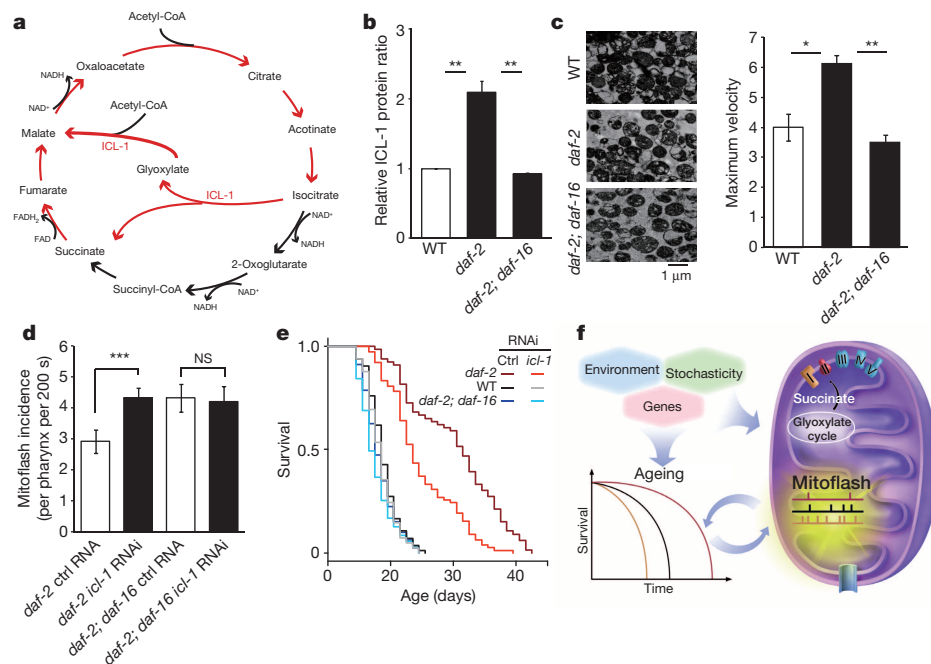


Figure 4 | Low mitoflash activity due to mitochondrial metabolic shift contributes to the longevity of the *daf-2* mutant. **a**, The tricarboxylic acid cycle and the glyoxylate cycle. **b**, Relative ICL-1 protein abundance in isolated mitochondria determined by ¹⁵N metabolic labelling and quantitative mass spectrometry. Ratios are (unlabelled ICL-1 of indicated strain/¹⁵N-labelled ICL-1) divided by (unlabelled ICL-1 of WT/¹⁵N-labelled ICL-1) and are shown as means ± s.e.m. (*n* = 3). **c**, Right: higher complex II activity (means ± s.e.m., *n* = 3) in the mitochondria isolated from the *daf-2*(*e1370*) mutant than in those purified from WT or the *daf-2*(*e1370*); *daf-16*(*mu86*) double mutant. Units of maximum velocity are 10⁻³ absorbance units at 600 nm per minute per microgram of protein. Left: electron micrographs of the mitochondrial samples used in the assay. **d**, Effect of *icl-1* RNAi on the mitoflash activity in *daf-2*(*e1370*) and the *daf-2*(*e1370*); *daf-16*(*mu86*) double mutant. Means ± s.e.m.,

n = 25 worms. **e**, *icl-1*(RNAi) shortened the lifespan of the *daf-2*(*e1370*) mutant but not the WT or the *daf-2*; *daf-16* double mutant. *P* < 0.001 between *daf-2* and *daf-2*; *icl-1*(RNAi), log-rank. *n* = 84/82, 66/79 and 80/71 for WT, *daf-2* and *daf-2*; *daf-16* animals on control/*icl-1* RNAi, respectively. In three independent experiments, *icl-1*(RNAi) significantly shortened the lifespan of the *daf-2*(*e1370*) mutant by 27% (shown), 12% and 15% (not shown). **f**, A model depicting the coupling between mitochondrial activity and ageing under the influence of genetic, environmental and stochastic factors. Note that a shift towards enhanced glyoxylate cycle and complex II activity decreases mitoflash production while extending the lifespan. Asterisk, 0.01 ≤ *P* < 0.05; two asterisks, 0.001 ≤ *P* < 0.01; three asterisks, *P* < 0.001; Student's *t*-test (**b**, **c**); Wilcoxon rank-sum test (**d**).

(Fig. 1c, d), it diminishes in mitochondria with severe damage²³. Among the day-9 adults, senile individuals with dysfunctional mitochondria and those that are still physiologically youthful and contain healthy mitochondria would all display low mitoflash activity, concealing the difference. In comparison, day 3 is early enough for all individuals to be young and healthy and late enough for various factors (genetic programs, past experience and random factors) to have had time to consolidate their effects.

Previously identified *C. elegans* ageing biomarkers²⁴ and lifespan predictors^{22,25–27} are effective only in mid-aged animals or require continual measurement over the course of one week. Some of them can account for lifespan variations attributed to stochastic factors, but none have a demonstrated predictive power for genetic and environmental factors. The day-3 mitoflash frequency is a unique lifespan predictor because it is an early, single-point measurement, yet is effective for all three classes of lifespan modifiers. This finding provides the most direct evidence for a central role of mitochondria in lifespan regulation.

On the basis of this observation, we further propose that mitochondrial activity is a determinant of lifespan. To test this we focused on the *daf-2*(*e1370*) mutant, whose remarkable longevity has catalysed research into the molecular mechanisms of ageing since 1993 (ref. 10). Isocitrate lyase 1 (ICL-1), a key enzyme of the glyoxylate cycle and a transcriptional target of DAF-16, is upregulated in *daf-2* mutants^{28,29}. Also known as GEX-interacting protein 7 (GEI-7), ICL-1 is a bifunctional enzyme (isocitrate lyase and malate synthase) that breaks isocitrate into succinate and glyoxylate and also converts glyoxylate into malate (Fig. 4a). By shunting the tricarboxylic acid cycle, the glyoxylate cycle produces less NADH (the substrate of respiratory complex I) and proportionally more succinate (the substrate of respiratory complex II), thus effectively

diverting more fuel to complex II. Consistent with previous findings, we detected a 95–130% abundance increase of ICL-1 in mitochondria isolated from the *daf-2*(*e1370*) day-1 adults compared with the WT or the *daf-2*; *daf-16* mutant (Fig. 4b). In addition, isolated *daf-2* mitochondria had a 50% increase in complex II activity (Fig. 4c). Because there is less electron leakage and hence less basal superoxide production at complex II than at complex I³⁰, we postulated that the coordinated elevation of ICL-1 and complex II activity decreases mitochondrial superoxide production and hence mitigates the triggering of mitoflashes in *daf-2* young adults (Fig. 2, day 1 and day 3). Indeed, RNAi of *icl-1* fully restored the day-3 mitoflash activity of the *daf-2*(*e1370*) mutant (Fig. 4d) and shortened its lifespan by up to 27% (Fig. 4e), consistent with the shortening of the *daf-2*(*mu150*) lifespan by *icl-1*/gei-7 RNAi²⁸. These results highlight the role of mitochondria in organismal ageing.

We have shown that mitoflash activity in young adulthood is a lifespan predictor in *C. elegans*. In long-lived *daf-2* animals, decreased mitochondrial production of superoxide through a shift in metabolic flux from complex I to complex II contributes to a slowing of organismal ageing. These data strongly support a coupling between the mitochondrial function and the process of ageing, and lend credence to ageing theories that place mitochondria at a nodal point of lifespan regulation (Fig. 4f). Another message stemming from these findings is that ageing is a programmed process: its overall rate, as set by genetic, environmental and stochastic factors, can be projected from mitochondrial activity in early adulthood.

METHODS SUMMARY

cpYFP was expressed in *C. elegans* under the promoter and the mitochondrial localization sequence of SDHB-1 (ref. 7), a respiratory complex II subunit on the matrix

side. Mitoflash activity in the anterior pharyngeal muscles was measured in live worms for 150–200 s, using a spinning-disk confocal microscope equipped with a 63×, 1.4 numerical aperture oil-immersion objective (UltraVIEW VOX; PerkinElmer) at 488 nm excitation, 500–550 nm emission, with 30% laser power. Images of 1,000 pixels × 1,000 pixels (1 pixel = 0.012 μm²) and 1 μm optical thickness were acquired at 1 frame s⁻¹ with a 430-ms exposure.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 24 May 2013; accepted 9 January 2014.

Published online 12 February 2014.

- Harman, D. The biologic clock: the mitochondria? *J. Am. Geriatr. Soc.* **20**, 145–147 (1972).
- Wang, W. *et al.* Superoxide flashes in single mitochondria. *Cell* **134**, 279–290 (2008).
- Fang, H. *et al.* Imaging superoxide flash and metabolism-coupled mitochondrial permeability transition in living animals. *Cell Res.* **21**, 1295–1304 (2011).
- Gavrilov, L. A. & Gavrilova, N. S. The quest for a general theory of aging and longevity. *Science Aging Knowledge Environ.* **2003** (28) RE5 (2003).
- Szewczyk, A. & Wojtczak, L. Mitochondria as a pharmacological target. *Pharmacol. Rev.* **54**, 101–127 (2002).
- Alexeyev, M. F. Is there more to aging than mitochondrial DNA and reactive oxygen species? *FEBS J.* **276**, 5768–5787 (2009).
- Lemire, B. D., Behrendt, M., DeCorby, A. & Gaskova, D. C. *elegans* longevity pathways converge to decrease mitochondrial membrane potential. *Mech. Ageing Dev.* **130**, 461–465 (2009).
- Kenyon, C. J. The genetics of ageing. *Nature* **464**, 504–512 (2010).
- Bulina, M. E. *et al.* A genetically encoded photosensitizer. *Nature Biotechnol.* **24**, 95–99 (2006).
- Kenyon, C. *et al.* A *C.elegans* mutant that lives twice as long as wild type. *Nature* **366**, 461–464 (1993).
- Yang, W. & Hekimi, S. A mitochondrial superoxide signal triggers increased longevity in *Caenorhabditis elegans*. *PLoS Biol.* **8**, e1000556 (2010).
- Alavez, S. *et al.* Amyloid-binding compounds maintain protein homeostasis during ageing and extend lifespan. *Nature* **472**, 226–229 (2011).
- Maier, W., Adilov, B., Regenass, M. & Alcedo, J. A neuromedin U receptor acts with the sensory system to modulate food type-dependent effects on *C.elegans* lifespan. *PLoS Biol.* **8**, e1000376 (2010).
- Cho, S. C. *et al.* DDS, 4,4'-diaminodiphenylsulfone, extends organismic lifespan. *Proc. Natl Acad. Sci. USA* **107**, 19326–19331 (2010).
- Liao, V. H. *et al.* Curcumin-mediated lifespan extension in *Caenorhabditis elegans*. *Mech. Ageing Dev.* **132**, 480–487 (2011).
- Robida-Stubbs, S. *et al.* TOR signaling and rapamycin influence longevity by regulating SKN-1/Nrf and DAF-16/FoxO. *Cell Metab.* **15**, 713–724 (2012).
- Choi, S. S. High glucose diets shorten lifespan of *Caenorhabditis elegans* via ectopic apoptosis induction. *Nutr. Res. Pract.* **5**, 214–218 (2011).
- Cypser, J. R., Tedesco, P. & Johnson, T. E. Hormesis and aging in *Caenorhabditis elegans*. *Exp. Gerontol.* **41**, 935–939 (2006).
- Honda, Y., Tanaka, M. & Honda, S. Trehalose extends longevity in the nematode *Caenorhabditis elegans*. *Aging Cell* **9**, 558–569 (2010).
- Panowski, S. H. *et al.* PHA-4/Foxa mediates diet-restriction-induced longevity of *C. elegans*. *Nature* **447**, 550–555 (2007).
- Steinbaugh, M. J., Sun, L. Y., Bartke, A. & Miller, R. A. Activation of genes involved in xenobiotic metabolism is a shared signature of mouse models with extended lifespan. *Am. J. Physiol. Endocrinol. Metab.* **303**, E488–E495 (2012).
- Pincus, Z., Smith-Vikos, T. & Slack, F. J. MicroRNA predictors of longevity in *Caenorhabditis elegans*. *PLoS Genet.* **7**, e1002306 (2011).
- Hou, Y. *et al.* Permeability transition pore-mediated mitochondrial superoxide flashes mediate an early inhibitory effect of Aβ_{1–42} on neural progenitor cell proliferation. *Neurobiol. Aging* <http://dx.doi.org/10.1016/j.neurobiolaging.2013.11.002> (18 November 2013).
- Herndon, L. A. *et al.* Stochastic and genetic factors influence tissue-specific decline in ageing *C. elegans*. *Nature* **419**, 808–814 (2002).
- Huang, C., Xiong, C. & Kornfeld, K. Measurements of age-related changes of physiological processes that predict lifespan of *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **101**, 8084–8089 (2004).
- Hsu, A. L., Feng, Z., Hsieh, M. Y. & Xu, X. Z. Identification by machine vision of the rate of motor activity decline as a lifespan predictor in *C. elegans*. *Neurobiol. Aging* **30**, 1498–1503 (2009).
- Sanchez-Blanco, A. & Kim, S. K. Variable pathogenicity determines individual lifespan in *Caenorhabditis elegans*. *PLoS Genet.* **7**, e1002047 (2011).
- Murphy, C. T. *et al.* Genes that act downstream of DAF-16 to influence the lifespan of *Caenorhabditis elegans*. *Nature* **424**, 277–283 (2003).
- Dong, M. Q. *et al.* Quantitative mass spectrometry identifies insulin signaling targets in *C. elegans*. *Science* **317**, 660–663 (2007).
- Yankovskaya, V. *et al.* Architecture of succinate dehydrogenase and reactive oxygen species generation. *Science* **299**, 700–704 (2003).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank B. D. Lemire for providing pSDH2(P-2)GFP; H. Jiang for providing KillerRed; the *Caenorhabditis* Genetics Center (funded by the National Institutes of Health Office of Research Infrastructure Programs (P40 OD010440)), D. Gems, P. Zhang and C. Kenyon for providing *C. elegans* strains; L. Miao, P. Liu, F. Yang, J. Li and J.-W. Lu for help with the electron microscopy and mitochondrial purification; L.-L. Du for suggestions; S. Sa, I. Bruce and X. Wang for critical reading of the manuscript; X. He for help with experiments; N. Yang for drawing the model in Fig. 4f; and S. Ji for discussion. This work was funded by the Ministry of Science and Technology of China (973 grants 2010CB835203 to M.-Q.D. and 2013CB531200 to H.C.), the National Natural Science Foundation of China (grants 31130067 and 31221002 to H.C.) and the municipal government of Beijing.

Author Contributions E.-Z.S. and C.-Q.S. designed experiments, generated strains, collected and analysed data and prepared figures. Y.L. collected and analysed data. W.-H.Z. and P.Z. performed lifespan analyses. P.-F.S. and Y.S. performed statistical analyses. W.-Y.L., J.X. and C.Z. processed image data and prepared figures or video files. N.L. helped to characterize the *C. elegans* mitoflash. X.W. was involved in the study design and manuscript writing. H.C. and M.-Q.D. designed the study, interpreted the data and wrote the paper.

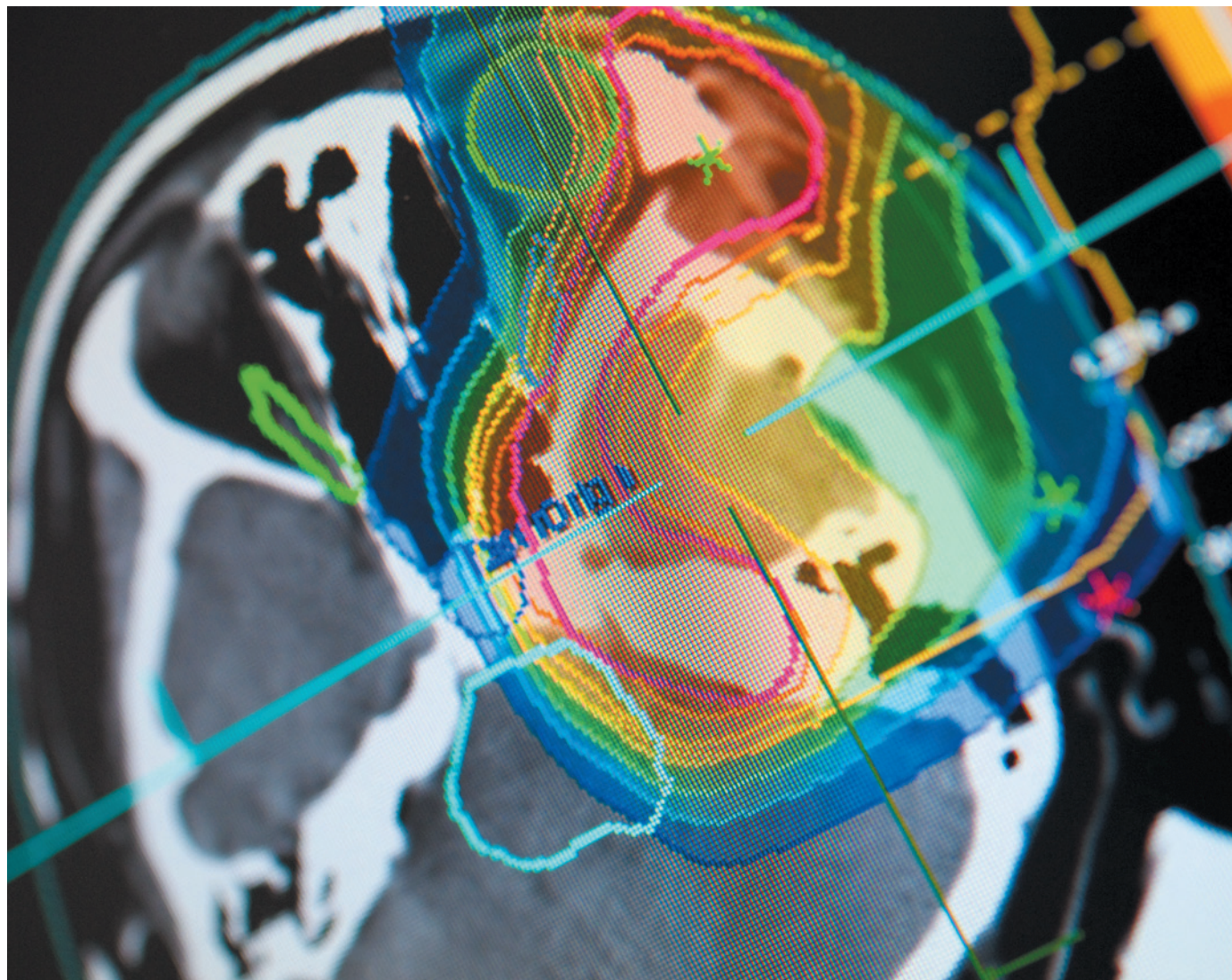
Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.-Q.D. (dongmengqiu@nibs.ac.cn) or H.C. (chengp@pku.edu.cn).

TECHNOLOGY FEATURE

SHARP SHOOTERS

Beams of charged particles can treat cancer more safely and effectively than X-rays. Physicists and biomedical researchers are working to refine the technology for wider use.

HEIDELBERG UNIV. HOSPITAL



A computed-tomography scan shows a tumour and the ion-beam dosage that will be used to treat it.

BY VIVIEN MARX

Clinicians attack cancer with many types of weapon, ranging from scalpels to physically remove all or most of a tumour to drugs that kill the tumour cells where they are. In about half of people with cancer, doctors go after the malignant cells with ionizing radiation.

Classic radiation treatment involves mainly X-rays. But because these lose energy all

along their path through the body — damaging healthy cells as they go — clinicians and researchers are increasingly paying attention to beams that use charged particles such as protons and carbon ions¹. Charged particles can deposit most of their lethal energy mainly at the tumour site, largely sparing the healthy tissue. Protons are slightly more lethal to cancer cells than X-rays, and carbon ions seem to be around 2–3 times as deadly.

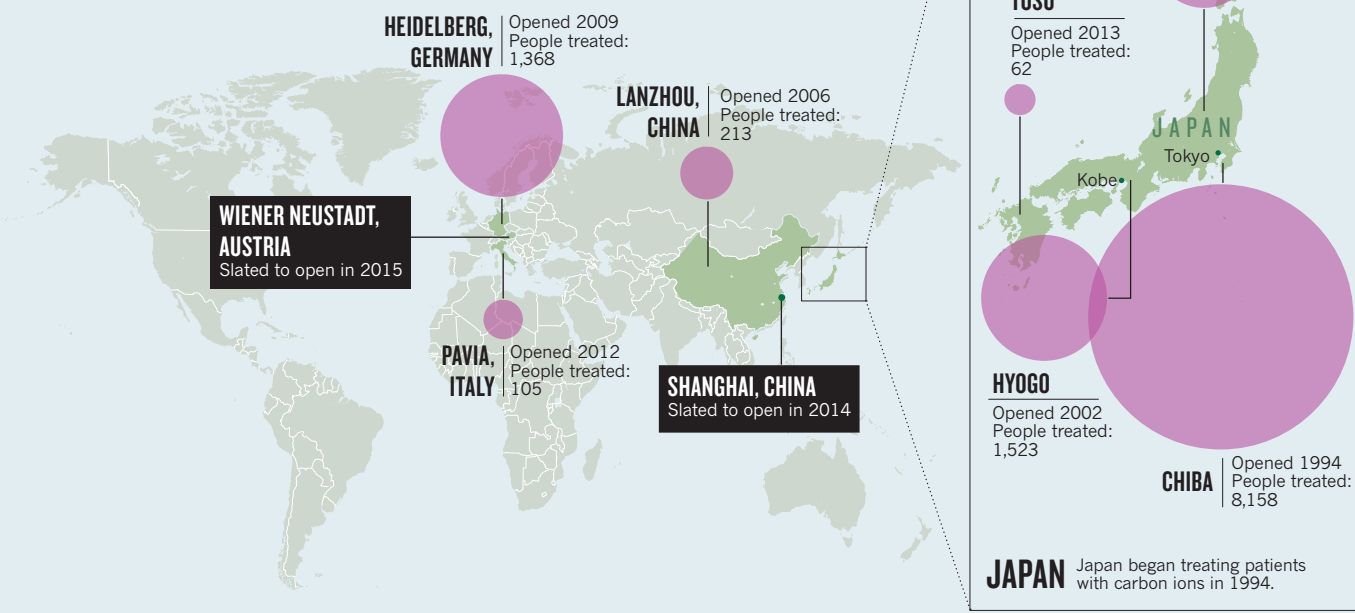
Worldwide, around 100,000 people have

received proton treatments for cancer. Japan, China, Germany and Italy have built ion-beam facilities that have treated some 12,000 patients with carbon ions, the majority in Japan and Germany (see ‘Carbon count’).

Carbon ions are heavier than protons, so the facilities to deliver them are pricier. The charged-particle facilities in Germany and Japan cost between US\$130 million and \$200 million each to build. Nonetheless, there has been a spike in research and ►

CARBON COUNT

Around 12,000 patients worldwide have been treated at dedicated carbon-ion facilities in Europe, China and Japan. The construction of two new facilities, encouraging clinical-trial results and advances in the technology mean those numbers are likely to grow.



► clinical activity to use charged particles more broadly for cancer treatment, and a hope that as the technology evolves, the price will come down.

Clinical trials and new types of radiobiology assay are under development to study the molecular effects of carbon ions as well as helium, lithium and oxygen ions, often referred to in cancer research as 'heavy' ions because they are heavier than protons. Advancing this approach to patient treatment will require further development of particle-accelerator and beam-delivery technology. An international community of clinicians, researchers and technology developers are working to make it happen.

The United States began treating patients with protons in the 1950s and has funded research in this area², but it has lagged behind Japan and Germany in advancing other ion therapies. The United States now has 14 facilities for proton treatment and none for carbon-ion treatment.

In Japan, radiation oncologist Hirohiko Tsujii of the National Institute of Radiological Sciences (NIRS) in Chiba has devoted his career to advancing carbon-ion therapy. Chiba is home to the first of Japan's four carbon-ion facilities, all of which the NIRS oversees, and treated its first patient in 1994. In Germany, the Heidelberg Ion-beam Therapy Center (HIT) has been treating patients since 2009 with protons and carbon ions³.

In both countries, the development and construction of these facilities was helped along by public funding for physics-based facilities

, as well as investments from large companies such as Siemens Healthcare, based in Erlangen, Germany, and Hitachi, Mitsubishi, Sumitomo and Toshiba, all based in Tokyo.

The published clinical results from patients in Germany and Japan are causing the United States to shift its stance, Tsujii says. The US National Cancer Institute (NCI) now wants to fund domestic research on carbon ions and other ion species, and to help fund an international clinical trial.

FIRE AWAY

Standard radiation treatment involves firing a barrage of X-rays at tumours in the hope of stunting their growth by breaking their cells' DNA. But some areas of tumours seem to be resistant to X-ray damage, and those that are not can often repair it. And many X-rays pass through tumour cells without hitting the DNA, leaving the cells unscathed. X-ray beams can also hit tissue adjacent to a tumour, which can set off molecular events leading to tumour formation. For example, children who survive cancer after radiation treatment face a heightened risk of secondary cancers later in life.

This kind of collateral damage has been reduced by advances in radiation oncology, such as the ability to deliver beams from different angles and with varying intensities so that they converge on the tumour. But improvement is still needed. For example, when treating head and neck cancers with conventional radiation, "we do a reasonable job", says Stephen Hahn, a radiation oncologist at the Perelman School of Medicine at the

University of Pennsylvania in Philadelphia — but the X-rays can harm healthy tissues nearby, such as the heart, oesophagus or lung.

Charged particles, by contrast, can mostly avoid healthy tissue. Stripped of their electrons and accelerated to some 70% of the speed of light, they pass through healthy tissue without interacting strongly with the atoms there. Once their speed drops to a certain level, however, they abruptly deposit almost all of their energy in "a dramatic bam", says Arnold Pomposh, a physicist at the University of Texas Southwestern Medical Center in Dallas.

This deposition of energy is known as the Bragg peak, named after physicist William Henry Bragg, who discovered the behaviour in 1903. Charged-particle beams can be tuned so that particles reach their Bragg peaks right at a tumour, where they do the maximum damage.

READY, AIM, SCAN

The researchers at the HIT use — and are continuing to develop — an approach called raster scanning, in which a source readies one or several types of ion that are then accelerated. Next, the beams are extracted slowly in bunches, each of which has hundreds of individual Bragg-peak positions tailored to deposit radiation throughout the tumour.

The position, size and intensity of these collections of beams are measured around 100,000 times a second to ensure safety and precision. The beams fill the tumour's contours as a hand fills a glove, says Thomas Haberer, chief technology officer of the HIT, who developed the technology while working at the GSI

Helmholtz Centre for Heavy Ion Research in Darmstadt, Germany. (The HIT is a spin-out of that centre.)

“Clinically, we use protons and carbon-ion radiation as is needed from one patient to the next,” says Haberer. He and his team are working to add oxygen, helium and other ions to the mix. The researchers also want to increase the number of ion species they use, because the heavier ones stay focused even when travelling deep into the body, as in the treatment of prostate cancer, he says. The heavier the ion, the greater the possible damage to solid tumours.

For tumours that contain a mix of healthy and diseased cells, the HIT team chooses protons to avoid injuring the healthy ones.

During particle-beam treatment, a patient lies on a couch wearing a stiff plastic garment that positions them with millimetre-scale precision in reference to a computed-tomography image of the tumour, says Haberer. A tumour can shift as the patient breathes, so the system must also adjust to account for this change.

In collaboration with several research groups, Haberer and his colleagues are working on ways to use the raster-scanner approach to refine how a tumour can be tracked during treatment. “In the lab, the prototype is working well,” he says. “It will take a while for it to reach clinical application, at least a few years.”

At the HIT, clinicians have applied carbon-ion therapy to tumours in the brain and at the base of the skull, and to head and neck cancers. And they have begun treating cancers of the liver and pancreas, recurrent rectal and prostate cancers, and paediatric bone cancer. Children are regularly treated with proton beams, and the team wants to launch clinical trials of carbon ions for paediatric cancers, says Haberer.

When treating cancers of organs in the torso, clinicians use ‘spacers’, which can be made from a variety of materials, to physically shift healthy tissue adjacent to the tumour out of the way of the Bragg peak. For example, Tsujii says, in carbon-ion therapy for colon cancer, “we put a spacer between the tumour and the intestines”.

Tsujii points to the expanding range of tumour types that have been treated in Japan with carbon ions. When people with rectal cancer are treated with surgery, he notes, around 15% develop recurrence within 3–5 years. Another surgery is an option for only 10–40% of them. When these patients



Gantries to direct charged-particle beams weigh as much as 600 tonnes.

are treated with carbon ions, only 10% of them develop a second recurrence, compared with 30–70% of those treated with X-rays. Patients eventually succumb to metastases, but Tsujii nonetheless finds the results so far promising.

Another study under way in Japan is looking at combined chemotherapy and carbon-ion therapy to treat people who have inoperable pancreatic tumours, and at pre-operative carbon-ion therapy for pancreatic cancer that can be removed surgically. Carbon-ion therapy may take less toll on the patient and reduce treatment times — for liver cancer and early-stage lung cancer, carbon ions are delivered in one or two sessions over as many days, compared with 10–30 sessions over many days or weeks for X-ray therapy.

GROWING THE TECHNOLOGY

Protons hinder the growth of some tumours better than X-rays, says Herman Suit, a proton-therapy pioneer at Massachusetts General Hospital in Boston⁴. Some of the published results from carbon-ion treatment centres in Japan and Germany are “impressive”,

he adds, such as those for tumours at the base of the skull, renal cancer and mucosal melanoma of the head and neck. He would like to see clinical trials comparing protons and carbon ions.

Carbon-ion therapy has so far been used mainly on tumours that are difficult to remove surgically and risky to treat with classic radiation. Tumours at the base of the skull, for example, are near nerves, brain tissue and the cochlea of the inner ear, where X-ray exposure could cause debilitating damage.

Hahn says that some charged particles deliver “a more powerful punch” to tumours and that he and most of his colleagues now accept that carbon-ion therapy performs well on the most challenging tumours. The next step is to see whether charged particles are right for more common diseases such as lung cancer.

COMPARING BEAMS

According to the NCI, the first international clinical trial on charged particles is now being planned. Slated to last 3–5 years, the randomized phase III trial, which will examine efficacy and dose, will compare X-rays, protons and carbon ions in the treatment of cancers of the pancreas, liver, head and neck, as well as bone and soft-tissue tumours and recurrent rectal cancer. The NCI will contribute funding for the trial and is soliciting proposals from US institutions and

from carbon-ion facilities in Japan, Germany, Italy and China.

The trial presents many logistical challenges in addition to that of obtaining the necessary scientific review and approval. Facilities for charged-particle therapy are much less common than those for X-ray treatment, so patients will need to travel further to reach them, especially the carbon-ion centres.

The NCI also wants to fund domestic research into charged-particle therapy. James Deye, a programme director for extramural radiation research at the NCI, and his colleagues are poring over project applications from institutions vying to set up the first US research centre for particle-beam radiation therapy. The NCI funds research, not construction, Deye says, so applicants must find the money to build, over the next 5–10 years, a research facility that can handle treatments with protons, carbon ions and other ion species. Grant recipients will be announced later this year.

One of the applicants is the University of Texas Southwestern Medical Center, where radiation oncologist Hak Choy wants to put

classic radiation, protons and heavy ions all under one roof. A new X-ray facility is set to open by 2016; a proton-therapy suite is slated for 2017; and Choy hopes to have a carbon-ion facility by 2021. The project is under review with the Texas state government, and Choy is hopeful about receiving support so that construction can begin.

ACCELERATE AND DELIVER

Accelerators and beam-delivery systems are crucial components in directing charged particles at patient tumours, but their size and operation costs are some of the factors hindering their widespread use. To help to make the use of this technology more feasible, some scientists are trying to make these systems more compact.

One overhaul researchers are targeting is of the gantry, a massive rotating platform that delivers ion beams to patients at any angle. To achieve different angles, the gantry directs the ions with huge, powerful magnets, which weigh it down and raise its electricity use. The gantry at the HIT is as tall as a commercial passenger aircraft and weighs around 600 tonnes. Because carbon ions have more momentum and charge than protons, a beam of them is around 2.5 times harder to bend than a proton beam, says Stephen Peggs, an accelerator physicist at Brookhaven National Laboratory in Upton, New York.

Using stronger magnets on the gantry would be beneficial, he says. Superconducting magnets could be an option, he adds, because they are not limited to a magnetic strength of 1 Tesla, as conventional magnets are, and could have strengths closer to 4 or 5 Tesla. That boost could reduce the gantry size, because the greater the force the magnet exerts on the beam, the more readily the beam can be bent in a smaller radius, reducing the size of the system.

“Compared to X-rays, charged particles deliver ‘a more powerful punch’ to tumours.”

Peggs and his team are also working to improve the accelerator itself. Charged-particle therapy accelerators could be more compact, he says, if they cycled more quickly.

The Brookhaven scientists are building a prototype of a rapid-cycling accelerator, which extracts beams more often than the slow-cycling accelerators now used in carbon-ion treatment facilities⁵. In the new system, fewer ions travel around the accelerator track at any given time, and these doses are extracted cyclically and delivered to the patient.

The use of fewer ions and more frequent extraction reduces the size of the beam pipe and the other system components, including the conventional magnets used on the accelerator to direct the beam. Such design changes also cut power use relative to that of slow-cycling systems.

The Brookhaven researchers are testing this concept by building a fast-cycling accelerator to deliver multiple ion species, including lithium, neon, helium and carbon. It will be able to deliver a salvo of one ion species and then quickly switch energy levels and deliver another, Peggs says.

The scientists are building the system components but want to assemble them at a location — not yet determined — where they can be applied to biomedicine. The prototype could be used as a radiobiology research facility to continue maturing the technology for charged-particle therapy, Peggs says.

To help with technology transfer, the Brookhaven scientists have partnered with Best Medical, a company in Springfield, Virginia, that builds radiation facilities. Krishnan Suthanthiran, the company's chief executive, sees a market for a rapid-cycling system.

His company has spent around \$5 million on the partnership thus far and expects it will take another \$10 million to \$15 million to build a working system by 2016. This year, he is applying for approval from the US Food and Drug Administration to build a facility. He estimates that a therapeutic centre will cost between



Stephen Peggs is building a fast-cycling accelerator.

\$30 million and \$100 million, depending on how many treatment rooms it has. It may or may not need a gantry and will probably require less radiation shielding than current carbon-ion facilities, reducing cost and footprint.

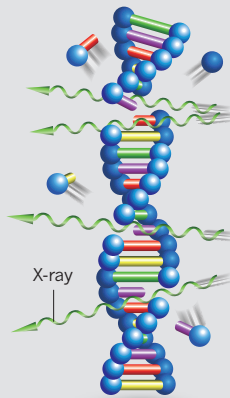
Haberer says that he likes Peggs' concept because of its robust, steady way of cycling, which is perhaps even more stable than that of a slow-cycling system. But until a rapid-cycling system is built and used in therapy, it will not be easy to compare systems, he says. Superconducting magnets would allow the facility's dimensions to be reduced, but they do not yet work quickly enough. "At present, these magnets are slow — which would mean fewer patients could be treated, impinging on the facilities' sustainability," Haberer says. And, he notes, a rapid-cycling accelerator is better able to extract the exact dose at the right time, but it could be difficult to monitor the beams extracted to ensure their dose and

GREATEST HITS

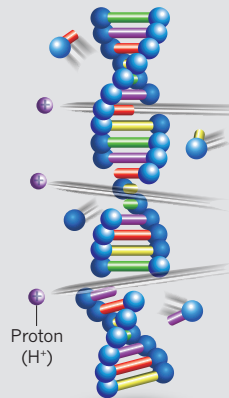
Radiation can kill cancer cells by damaging their DNA. X-rays can hit or miss. Protons are slightly more lethal to cancer cells than X-rays. Carbon ions are around 2–3 times as damaging as X-rays.



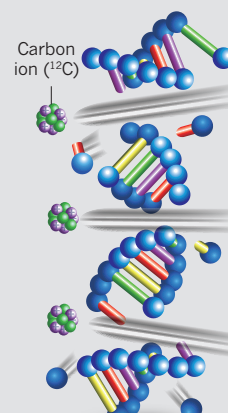
DNA



X-ray



Proton beam



Carbon-ion beam



HEIDELBERG UNIV. HOSPITAL

The treatment area of an ion-beam facility (left). Patients are carefully positioned (right) to ensure that the particle beam hits the tumour accurately.

directionality.

To further explore the potential uses and ideal dosages for charged-particle therapy, scientists want to expand basic radiobiology research. Scientists perform this work at charged-particle therapy centres and in facilities such as the NASA Space Radiation Biology Lab at Brookhaven, for example.

Kathryn Held, a radiobiologist at Massachusetts General Hospital in Boston is working on ways to get molecular information about the effects of different ion species on tumour cells — information that could help researchers to develop dosage regimens.

New approaches to studying cell survival, changes in cell cycle and cell death will help researchers to explore why charged particles have more tumour-killing power than X-rays.

Measures such as standard 'clonogenic assays', in which scientists irradiate cells and then see whether they continue to grow and form colonies, suggest that protons are slightly more effective than X-rays at killing cells, whereas carbon ions are about 2–3 times more effective. Scientists have a number of hunches about why this is so.

One is related to the fact that X-rays tend to be spread out, meaning that many of them pass through the cell without hitting DNA, says Held. Ions such as carbon are heavier and bigger than protons, have a greater positive charge and move more slowly through tissue

or cells, creating a thicker track of ionization. This track seems to produce clusters of damage, such as breaks in one or both strands of DNA, and damage to neighbouring nucleotides (see 'Greatest hits').

When one strand of DNA breaks, repair enzymes use the sequence of the other, intact strand as a template to fix the helix's rails and rungs. But double-strand breaks are harder to repair accurately, because there is no intact template from which to reconstitute the DNA.

"Assays that specifically measure types of possible clustered DNA damages would be very useful."

antibodies to detect DNA-repair enzymes — helps to explain the type of injury that charged particles inflict on tumour cells. "However, we do not have good assays to quantify, or identify the composition of, those clustered lesions," Held says. "DNA assays that more specifically measure the various types of possible clustered DNA damages would be very useful."

Another reason charged particles may pack more punch to cancerous cells is related to tumour physiology. As tumours grow, oxygen-poor regions develop, and these areas seem to

A cluster of damage adds to the challenge. All of this helps to explain why carbon ions are more lethal, says Held.

Indirect evidence — computer simulations or the use of

be resistant to classic radiation treatment. That is because X-rays kill cells by producing free radicals — reactive molecules formed from the water in cells and tissues — which then react with DNA to produce other destructive radicals. More oxygen exacerbates the damage these radicals can cause, and less oxygen weakens their effect. Researchers think that charged particles such as carbon ions may not lose their destructive power in low-oxygen regions.

Basic radiobiology research on charged particles can feed into clinical practices in current and future facilities. Radiation oncologist Anders Brahme at the Karolinska Institute in Stockholm, who has spent his career working on charged-particle therapy, says that he sees great potential for cancer treatment in such therapies and is excited that they are drawing global interest. In his view, charged particles offer the chance for radiation oncology to move from a cancer treatment to a cancer cure. ■

Vivien Marx is technology editor for *Nature* and *Nature Methods*.

1. Durante, M. & Loeffler, J. S. *Nature Rev. Clin. Oncol.* **7**, 37–43 (2010).
2. Deye, J. A. *Health Phys.* **103**, 662–666 (2012).
3. Tsujii, H. & Kamada, T. *Jpn J. Clin. Oncol.* **42**, 670–685 (2012).
4. Suit, H. *et al. Acta Oncol.* **42**, 800–808 (2003).
5. Peggs, S. *et al. in Proc. 8th Eur. Part. Accel. Conf.* 2754–2756 (EPAC, 2002).

natureOUTLOOK

SCHIZOPHRENIA

3 April 2014 / Vol 508 / Issue No. 7494



Cover art: Sue Morgan

Editorial

Herb Brody, Michelle Grayson, Tony Scully, Nick Haines, Victoria Kitchener

Art & Design

Wes Fernandes, Mohamed Ashour, Amr Rahma, Alisdair Macdonald, Alison Reed, Andrea Duffy

Production

Donald McDonald, Ian Pope, Christopher Clough

Sponsorship

Will Piper, Yvette Smith, Reya Silao

Marketing

Elena Woodstock, Steven Hurst

Project Manager

Christian Manco

Art Director

Kelly Buckheit Krause

Publisher

Richard Hughes

Magazine Editor

Rosie Mestel

Editor-in-Chief

Philip Campbell

The cover of this Outlook vividly conveys the sense of fragmentation experienced by many people with schizophrenia. The artist, Sue Morgan, has been living with the condition for about twenty years, although she is now in remission. Morgan tried in this illustration to “convey the sense of horrible overwhelmingness” of living with this condition, which still confounds researchers trying to both treat the symptoms and fix the underlying biology. The lines connecting to the head, she says, “express the invisible signals and connections that I can tap into.”

Science is finally beginning to offer slivers of hope. Studies of the language difficulties faced by schizophrenia patients are revealing clues to how the disorder affects the brain (page S4). Researchers have identified hundreds of genes associated with the disease (S6) — although these findings are creating new puzzles as they solve old ones (S20).

Frustratingly little progress has been made in drug development for twenty years, raising questions about the best research strategy (S18). Help could come in the form of better animal models (S8). And treatments may be close for the so-called negative symptoms of schizophrenia — the emotional withdrawal that can be as disabling as the psychosis more often associated with the disease (S10).

The disorder typically strikes with full fury in young adults, but warning signs are often evident before then, offering the possibility of earlier intervention (S12). And lessons might be learned from the developing world, where, paradoxically, people with schizophrenia tend to have better outcomes than in richer countries (S14). Scientists are also figuring out why the disease is associated with accelerated ageing (S16).

This Outlook was produced with support of a grant from Otsuka Pharmaceutical Development and Commercialization, Inc. As always, Nature Publishing Group retains sole responsibility for editorial content.

Herb Brody

Supplements Editor

CONTENTS

S2 AETIOLOGY

Searching for schizophrenia's roots

Brain chemistry may yield new treatments

S4 LANGUAGE

Lost in translation

Can speech problems tell us what's wrong?

S6 GENETICS

Unravelling complexity

From genomes to drug targets

S8 DRUG DEVELOPMENT

The modelling challenge

Animals help the search for therapies

S10 THERAPEUTICS

Negative feedback

Looking for the feel-good factor

S12 PREVENTION

Before the break

The benefits of spotting the early signs

S14 DEVELOPING COUNTRIES

The outcomes paradox

Can the West learn from poorer nations?

S16 AGEING

Live faster, die younger

Getting old before their time

S18 PERSPECTIVE

Retreat from the radical

Research strategy reforms go too far, says Stephen R. Marder

S20 PERSPECTIVE

Revealing molecular secrets

Steven E. Hyman ponders the growing challenge of schizophrenia genetics

COLLECTION

S21 Diverse types of genetic variation converge on functional gene networks involved in schizophrenia
Sarah R. Gilman et al.

S27 Mosaic copy number variation in schizophrenia
Douglas M. Ruderfer et al.

S32 A Randomized Exploratory Trial of an Alpha-7 Nicotinic Receptor Agonist (TC-5619) for Cognitive Enhancement in Schizophrenia
Jeffrey A. Lieberman et al.

S40 Schizophrenia as a pathway disease
Patrick F. Sullivan

S41 Schizophrenia, social environment and the brain
Heike Tost & Andreas Meyer-Lindenberg

S44 Genome-wide significant associations in schizophrenia to *ITIH3/4*, *CACNA1C* and *SDCCAG8*, and extensive replication of associations reported by the Schizophrenia PGC
M. L. Hamshere et al.

Nature Outlooks are sponsored supplements that aim to stimulate interest and debate around a subject of interest to the sponsor, while satisfying the editorial values of *Nature* and our readers' expectations. The boundaries of sponsor involvement are clearly delineated in the Nature Outlook Editorial guidelines available at http://www.nature.com/advertising/resources/pdf/outlook_guidelines.pdf

CITING THE OUTLOOK

Cite as a supplement to *Nature*, for example, *Nature* Vol XXX, No. XXXX Suppl. Sxx–Sxx (2014). To cite previously published articles from the collection, please use the original citation, which can be found at the start of each article.

VISIT THE OUTLOOK ONLINE

The *Nature Outlook Schizophrenia* supplement can be found at <http://www.nature.com/nature/outlook/schizophrenia/>

All featured articles will be freely available for 6 months.

SUBSCRIPTIONS AND CUSTOMER SERVICES

For UK/Europe (excluding Japan): Nature Publishing Group, Subscriptions, Brunel Road, Basingstoke, Hants, RG21 6XS, UK. Tel: +44 (0) 1256 329242. Subscriptions and customer services for Americas — including Canada, Latin America and the Caribbean: Nature Publishing Group, 75 Varick St, 9th floor, New York, NY 10013-1917, USA. Tel: +1 866 363 7860 (US/Canada) or +1 212 726 9223 (outside US/Canada). Japan/China/Korea: Nature Publishing Group — Asia-Pacific, Chiyoda Building 5-6th Floor, 2-37 Ichigaya Tamachi, Shinjuku-ku, Tokyo, 162-0843, Japan. Tel: +81 3 3267 8751.

CUSTOMER SERVICES

Feedback@nature.com
Copyright © 2014 Nature Publishing Group

SEARCHING FOR SCHIZOPHRENIA'S ROOTS

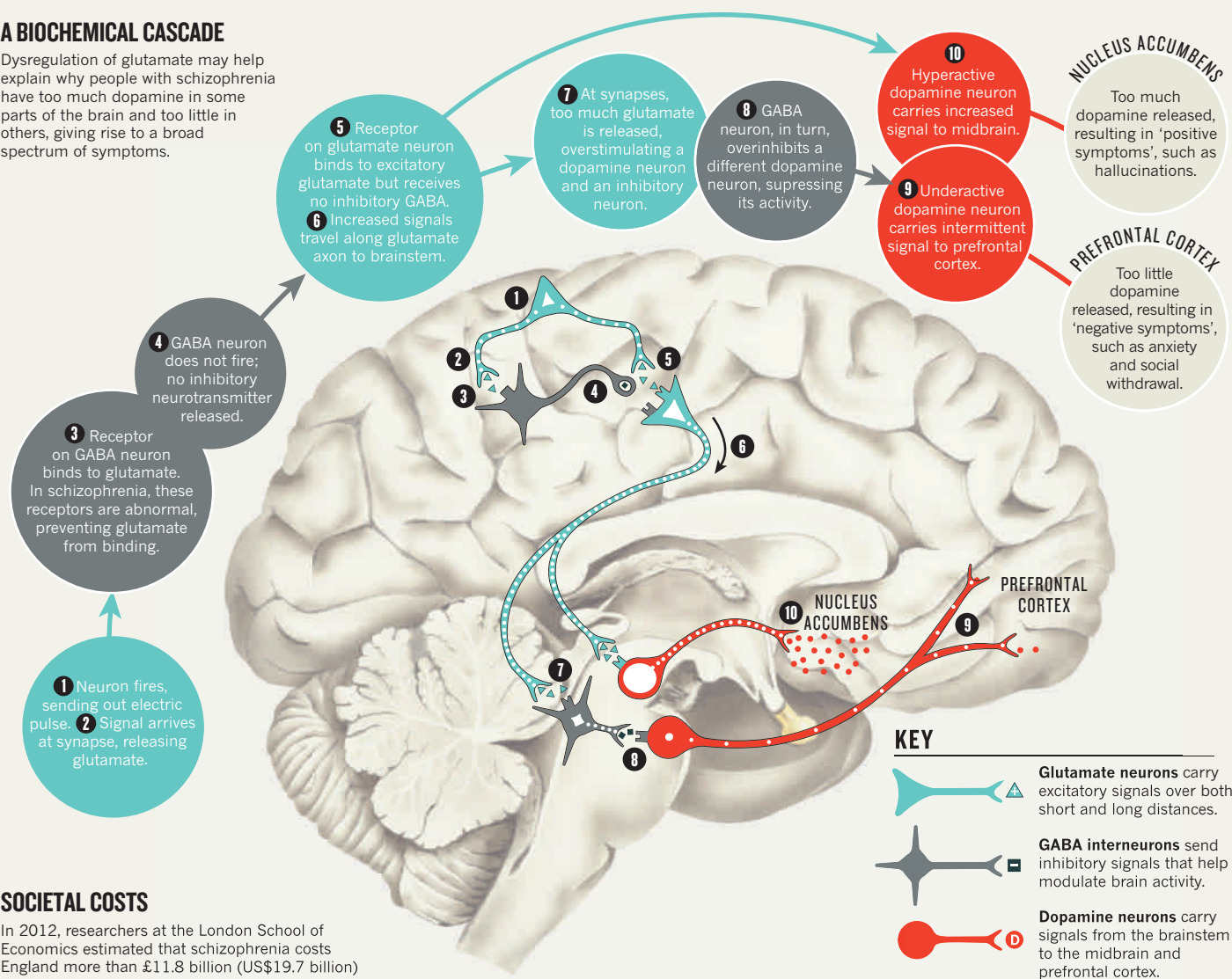
Sixty years after the first schizophrenia drug hit the market, researchers are still struggling to understand and treat the disorder. By **Emily Elert**.

1%

of adults have schizophrenia.

A BIOCHEMICAL CASCADE

Dysregulation of glutamate may help explain why people with schizophrenia have too much dopamine in some parts of the brain and too little in others, giving rise to a broad spectrum of symptoms.



SOCIETAL COSTS

In 2012, researchers at the London School of Economics estimated that schizophrenia costs England more than £11.8 billion (US\$19.7 billion) each year — nearly £76,000 for each person afflicted.

Premature mortality: Mortality rates are higher in people with schizophrenia, due to more suicides and poorer health.

Excess unemployment: One estimate found that just 10% of people with schizophrenia are employed, 60% below the UK average.

Institutional: Most people with schizophrenia have at least one in-patient stay and a high risk of readmission.



50%

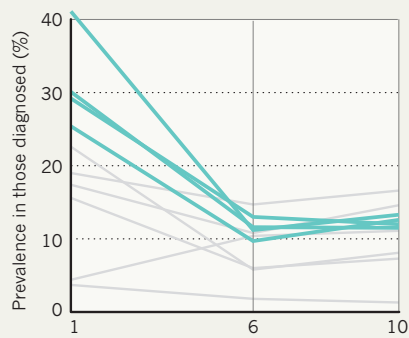
of people with schizophrenia are not receiving proper treatment.

THE CHANGING FACE OF SCHIZOPHRENIA

The symptoms of schizophrenia vary from person to person, but follow a characteristic pattern over the course of the disease.

POSITIVE SYMPTOMS

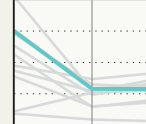
The onset of schizophrenia is manifested by symptoms associated with psychosis. These often respond well to antipsychotic drugs.



Hallucinations



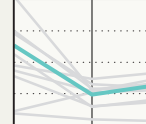
Thought disorder



Motor problems



Delusions

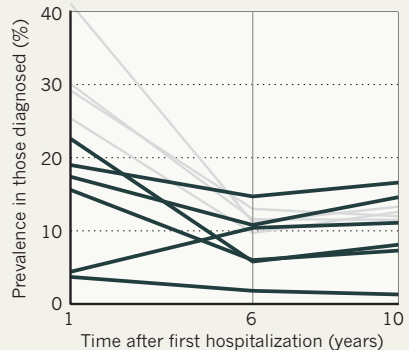


5.6%

is the suicide rate for people with schizophrenia.

NEGATIVE SYMPTOMS

These traits often linger or worsen, accompanied by impairments to cognitive processes, such as working memory.



Flat affect



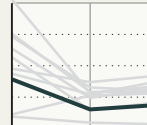
Odd behaviour



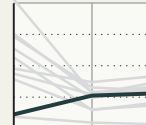
Social withdrawal



Apathy



Self-neglect

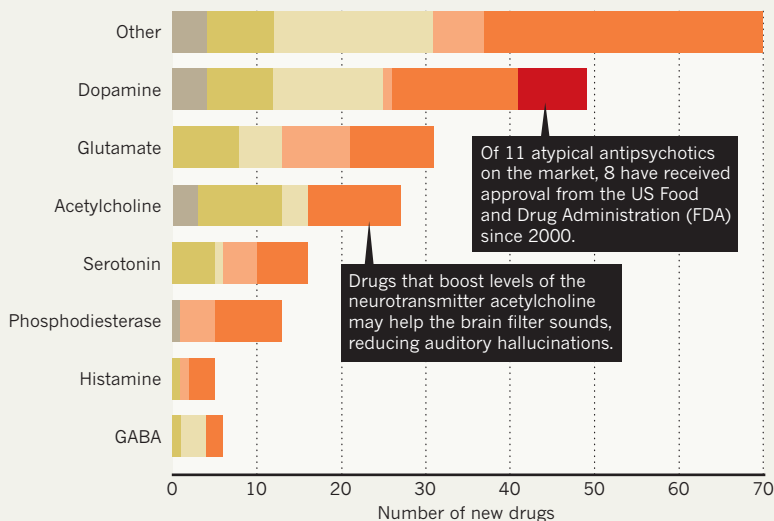


Anxiety



15 YEARS OF DRUG DEVELOPMENT

The most widely used schizophrenia drugs, atypical antipsychotics, are effective in treating positive symptoms. But researchers are finding new neural targets to address negative and cognitive symptoms.

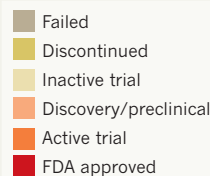


219

drugs have been tested for the treatment of schizophrenia since 1999.

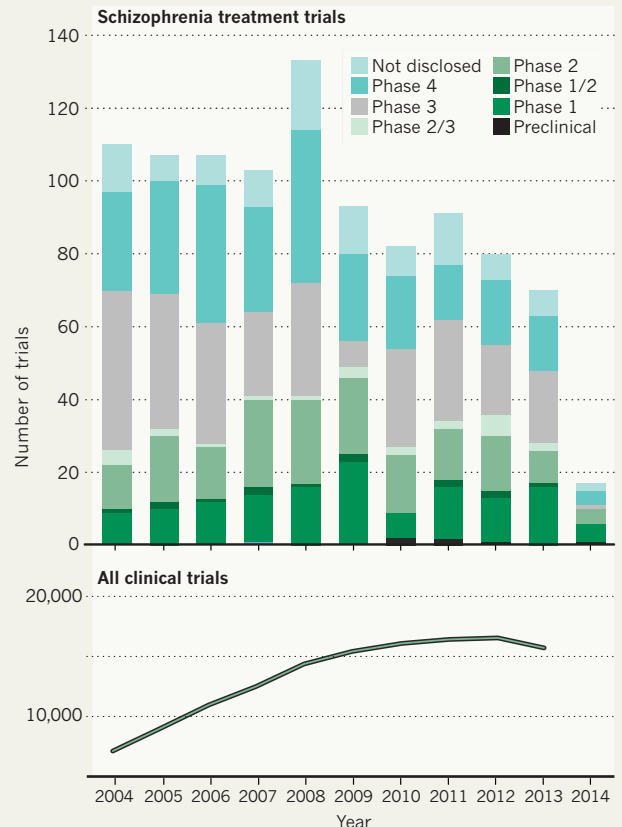
8

have received FDA approval.



SCHIZOPHRENIA SLOWDOWN

For the past ten years, clinical trials of schizophrenia treatments have been on a downward trajectory, even though clinical trials in general have been rising.





Gina Kuperberg (right) studies language processing by recording brain activity in reading tasks.

LANGUAGE

Lost in translation

Unravelling the mystery of verbal dysfunction in schizophrenia could yield clues to the nature of the disease.

BY DAVID NOONAN

The Swiss psychiatrist Eugen Bleuler is perhaps best known for coining the term ‘schizophrenia’ more than a century ago. But he also paved the way to understanding the disordered language that characterizes schizophrenia. Certain quotes from patients have since been enshrined in the medical literature as classic examples of the types of language dysfunction that typify schizophrenia. Two quotes continue to echo through time

decades after being uttered by patients. The first is from a 1911 book by Bleuler:

“I always liked geography. My last teacher in that subject was Professor August A. He was a man with black eyes. I also like black eyes. There are also blue eyes and gray eyes and other sorts, too. I have heard it said that snakes have green eyes. All people have eyes.”

The second, which came in response to a question about “current political issues, like the

energy crisis”, is from a 1986 paper by Nancy C. Andreasen:

“They’re destroying too many cattle and oil just to make soap. If we need soap when you can jump into a pool of water, and then when you go to buy your gasoline, my folks always thought they should, get pop but the best thing to get, is motor oil, and, money.”

In the first quote, Bleuler’s patient demonstrates ‘derailment’, in which the speaker loses focus and slides off topic — in this case he was suddenly distracted by the more intriguing subject of eyes. Andreasen’s patient exhibits a more profound disconnect: the inability to produce intelligible speech when prompted, instead tossing together a word salad that includes a few potentially relevant terms but carries no meaning.

Both these haunting passages are manifestations of thought disorder, and both have been evoked again and again by generations of researchers seeking to better understand the central role that language plays in schizophrenia. They endure because, in a strange and compelling way, they embody the entwined mysteries of language and thought.

“Language is like a window into the mind,” says cognitive neuroscientist Debra Titone at McGill University in Montreal, Canada. After decades of study, the rules and mechanisms of normal language function are well understood. We don’t know everything, but we know enough — and have developed enough research tools — to track the flow of language through the normal brain as words are chosen and arranged in sentences that make sense and carry meaning.

Investigators such as Titone use language to explore the brains and disordered minds of people with schizophrenia. Their goal is to identify and understand the underlying causes of the language dysfunction in schizophrenia, and to help determine the neural basis of this devastating mental illness. For patients, there is hope that this work will ultimately lead to earlier diagnosis, better treatments, and interventions that target specific brain functions and cognitive behaviours to improve language and quality of life.

Titone and others who study the language pathology associated with mental disorders make no distinction between language and thought. The once robust debate over that issue is long over. “Many of the stored representations that you are drawing on in thought are exactly the same as those that you are drawing on in language,” says Gina Kuperberg, a cognitive neuroscientist at Tufts University in

Medford, Massachusetts, who has applied psycholinguistics to the study of schizophrenia. Titone concurs: “If you look at what the brain is doing

➔ NATURE.COM
For recent research on language dysfunction:
go.nature.com/xhrVg5

when it's doing language, it's not as if there's some language organ that lights up."

MUDDLING THROUGH

Like the Bleuler and Andreasen examples (both cited by Kuperberg in a 2010 paper), schizophrenic discourse often retains the normal structure of language. It generally follows the rules of grammar: the basics are there, the phonology and even some of the syntax, but it is disjointed and hard to follow. Titone attributes this phenomenon to the fact that the structural properties of language are "drilled into our brains" from the moment we are born. By late adolescence or early adulthood, the typical age of onset for schizophrenia, it's already locked in.

In the disrupted communication that is a hallmark of schizophrenia, the main problem is how words are combined to produce overall meaning. The underlying issues involve the storage of meaning in the brain and the way individuals mobilize and update that information. "Meaning is the brain," says Kuperberg, who is also a psychiatrist at Harvard Medical School. "It is stored and represented extraordinarily widely throughout the brain." These mental representations of our external and internal worlds exist in complex neural networks that we draw on when we use language.

During conversation, mentally healthy people sort through this massive store of ever-changing information to produce comprehensible language. But in some people with schizophrenia this basic ability is impaired. The problem is often described as a "loosening of associations" in which, for example, words are chosen not because they make sense, but because the patient simply associates them with other words. When the man who professed his fondness for geography as it was taught by Professor A accessed his stored knowledge of the experience, he quickly veered off track and abandoned the professor altogether for a string of loosely connected references to eyes. For Andreasen's more severely impaired patient, the question about the energy crisis sent him bouncing around his brain like a pinball, lighting up stored networks that, when transmitted as language, amounted to gibberish.

People without schizophrenia update and reorganize this stored information from moment to moment as they interact with others and the external world. As they navigate through their day, they process everything from the identity of the person they are talking to (their boss, perhaps) to changes in the tone of voice. But Kuperberg's research suggests that people with schizophrenia have difficulty performing this moment-to-moment updating, limiting their ability to revise their stored knowledge and adjust to the flow of everyday communication. Such abnormalities have been linked to a breakdown in the fast, continuous interaction between prefrontal and temporal areas of the brain. Among other problems, this

can result in a kind of cognitive rigidity, sticking to a topic at the expense of coherence — for example, talking about the weather when the pharmacist asks about a prescription because someone on the bus talked about the weather a few minutes before. Kuperberg adds that people with schizophrenia are often unaware of their difficulties with communication.

MAKING SENSE

It is generally thought that people with schizophrenia have problems in the way they access and mobilize stored meaning, and not in the way the meaning is stored in the first place, but Titone raises another possibility. Stored meaning is based on experience of the real world, she says, and "people with schizophrenia have a different experience with the world", so perhaps the way meaning is stored — the structure of it — might play a contributing role.

We know that the supple nature of language itself, with its many nuances and ambiguities, can cause problems for people with schizophrenia. In one study, Kuperberg, who uses functional magnetic resonance imaging, electroencephalography and other methods to study language function in normal and abnormal subjects, presented her subjects with the following sentence: "Every morning at breakfast the eggs would eat." It's a crafty little construction in which the individual words are semantically related but the whole doesn't add up. During experiments that measured the brain's electrical activity, it fooled even healthy brains for a few hundredths of a second before the conflict with stored knowledge (the implausibility of the sentence) was recognized and resolved by additional processing. Unlike the control subjects, however, those with schizophrenia did not do the extra processing and

Language dysfunction in schizophrenia is a perfect storm of complexity.

did not register the sentence as implausible. Their ability to do the extra work to recognize and resolve the conflict was impaired. The patterns of two brain waves measured and recorded during the experiment — one sensitive to syntax and one to stored knowledge — suggested that, in the schizophrenia patients, an over-reliance on stored knowledge trumped the rules.

"Language is inherently ambiguous," says Titone. "Words point to two different meanings all the time. It takes a lot more cognitive machinery to be able to reduce that kind of ambiguity in the moment as we're processing or producing language." People with schizophrenia, she says, find it particularly difficult to cope with this.

Insensitivity to context ("the eggs would eat") is a factor not only in comprehension, but also in the way language is produced. In normal subjects, context serves as a guide when recruiting neural networks — it's how you distinguish

between various meanings of the word 'bear', for example. In people with schizophrenia, a reduced ability to use context helps foster loose associations. "If you have an incoming signal that triggers, in some very knee-jerk, reflexive way, some thought, it's going to be really hard for context to trump that," she says.

Another complication for some people with schizophrenia, especially with respect to comprehension, is the fact that humans process language word-by-word. This incremental processing is very fast, of course, and the normal brain must use the context it has already built up to process each new word it encounters in real time. "To do that," says Kuperberg, "we need to mobilize all our stored world knowledge really quickly, use it to combine individual words together to generate new meaning, and then update our real-world knowledge memory stores with this new meaning." In other words, it's the sort of higher-order language function that flummoxes people with the impaired semantic memory, working memory and context insensitivity found in schizophrenia.

LINGUISTIC ISOLATION

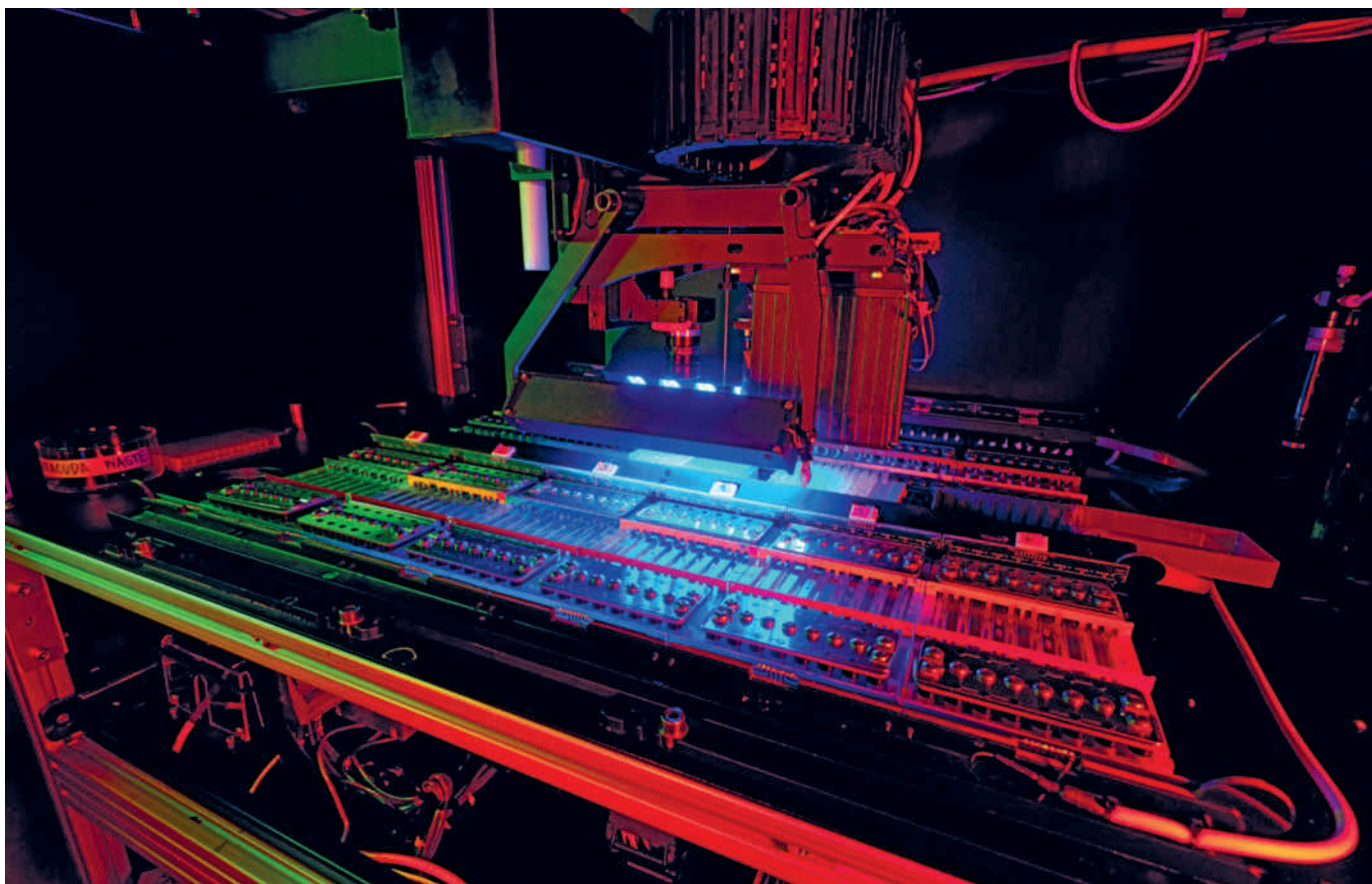
Kim Mueser, a psychologist at Boston University in Massachusetts who studies social skills in people with schizophrenia, points out yet another problem facing the sub-group of people with schizophrenia who demonstrate 'blunted affect'. These individuals seem emotionally flat but actually experience a full range of emotions — it's their ability to communicate them that is impaired. "When you talk about social skills," says Mueser, "90 per cent of what people are usually talking about are the verbal skills."

So, language dysfunction in schizophrenia is a perfect storm of complexity, bringing together a dauntingly complicated mental illness and an equally complicated higher cortical function. There is something especially cruel in the way schizophrenia cuts people off from the simple pleasures of conversation and denies them the comfort of sharing their thoughts and feelings with others.

Kuperberg hints at this idea when she departs from her scientific discourse to expound excitedly on the everyday miracle of two people talking. "You have your world and your agenda and your goals," she says, "and I have my world and my agenda and my goals, and we don't know each other, and yet, somehow, through this magical thing, this sequence of words, these syllables, somehow you are communicating to me through this code to somehow update my brain with your thoughts. Isn't that cool?"

Ordered thought, manifest in the beauty and wonder of language, is cool indeed. So too is the continuing effort of scientists to understand how that precious gift is lost in people with schizophrenia — and maybe one day to give it back to them. ■

David Noonan is a freelance science writer based in New Jersey.



MARTIN KRZYWINSKI/SCIENCE PHOTO LIBRARY

The use of genome sequencers has allowed researchers to identify many targets for possible schizophrenia drugs.

GENETICS

Unravelling complexity

A massive research collaboration is revealing hundreds of genes underlying schizophrenia risk, and may point the way to targeted treatments.

BY JESSICA WRIGHT

A plot of the genes associated with a psychiatric disease isn't the kind of image you'd normally find on a T-shirt. But at the 2013 Symposium on Emerging Genetics and Neurobiology of Severe Mental Illness, held in September at the Broad Institute in Cambridge, Massachusetts, more than 250 attendees snatched up garments displaying just such an image, along with the catchphrase "Schizophrenia genetics — a reality at last!" It's a clear sign that schizophrenia researchers are excited about recent advances in pinpointing the genetic causes of the disease.

Seven years ago, they had all but given up on this type of genetic analysis, which links single nucleotide changes in a gene with the likelihood of developing a disorder. Genome-wide association studies (GWAS) were failing to identify schizophrenia genes with statistical

certainty. But in the past two years, researchers have identified more than 100 genome-wide significant hits, implicating almost 700 genes. That tally may soon grow into the thousands. This remarkable progress has rekindled hope that researchers may be able to use genetics to identify targets for schizophrenia drugs.

Studies on twins have shown that as much as 80% of the risk of developing schizophrenia comes from the genes. But pinpointing targets for drug developers has proved an elusive goal. As a result, most schizophrenia drugs treat only the overt symptoms — the imaginary voices and other elements of psychosis — and do little to address the broad effects on the brain. They also have severe side effects. "Pharmaceutical companies have almost given up on the central nervous system entirely," says Steven Hyman, director of the Broad Institute's Stanley Center for Psychiatric Research. But now, he says, this may be about to change.

Much of the progress comes from technologies that allow researchers to scan large swaths of the genome to reveal unexpected genetic links, instead of looking only at individual genes. In addition, schizophrenia researchers have turned to the methods of 'big science', banding together to generate statistical power.

The driver behind this effort is the Psychiatric Genomics Consortium (PGC), which pools genetic data from researchers at 60 institutions in 19 countries. The PGC is collecting genomes focusing on five psychiatric disorders, including autism and attention deficit hyperactivity disorder (ADHD). But the number of genomes collected for schizophrenia surpasses the others: nearly 40,000 at a recent count.

The high level of collaboration for schizophrenia reflects the fact that most participants know how hard it is to unravel this complexity alone, says Patrick Sullivan, a geneticist at the University of North Carolina, Chapel Hill, and

the PGC's lead investigator. "We had to come together as a field."

The early findings reveal that schizophrenia results from mutations in multiple genes working together in different combinations. As researchers unearth more gene candidates, they are discovering biological pathways that may serve as entry points for drug therapies. "The scale of the polygenicity is surprising," says Steven McCarroll, a geneticist at Harvard Medical School in Boston, Massachusetts. "But the exciting thing is that these genetic hits aren't just random hits on the dashboard of the genome. They are aggregating in specific molecular complexes and pathways."

COMMON AND RARE

The first hits identified in schizophrenia were so-called rare variants — genetic mutations that usually have strong effects on their corresponding proteins. These mutations, which nearly always lead to symptoms and are fairly easy to identify, are often found in an affected individual but not in their unaffected parents or siblings. Studies suggest that rare variants account for up to 20% of schizophrenia cases. The remaining genetic hits are likely to be common variants, which are found throughout the population and have relatively mild effects, but in combination may be sufficient to tip the scales towards schizophrenia. Individuals may also have a combination of both variants.

It makes sense that variants with mild effects may be important, as many are in genes that influence crucial pathways, such as the signalling junctions between neurons. But schizophrenia is not characterized by the severe intellectual disability and epilepsy that is a hallmark of disrupted brain pathways. In fact, people with schizophrenia tend not to develop problems until adolescence. "The synapse is a hugely complicated machine — how can we mess with it in such a way that we can get a disorder where people develop normally for a couple of decades?" asks Sullivan.

Until recently, researchers didn't have the statistical power needed to test common variants' influence in schizophrenia. But the sheer number of participants gathered through the PGC allows researchers to pinpoint individual variants that contribute to schizophrenia risk.

In a 2013 *Nature Genetics* study, Stephan Ripke, a geneticist at the Broad Institute, and colleagues analysed the genomes of more than 21,000 individuals with schizophrenia¹. They identified 22 variants associated with schizophrenia at the level of genome-wide significance. At the Broad conference, Ripke presented — and wore — the latest data from this project: an analysis of 35,000 people with schizophrenia and 47,000 controls that has yielded more than 100 statistically significant hits implicating 672 genes. In the next year, the consortium aims to have analysed 100,000 genomes.

One of the new hits is in the gene *DRD2*, which makes a receptor for dopamine, a

chemical messenger that helps mediate reward. The discovery is not surprising, as all the existing drug therapies for schizophrenia target the dopamine pathway. But it is a proof of principle that the PGC can identify useful drug targets. *DRD2* is a "home run example", says Ripke. "It is very valid to assume that among the other hits there are other drug targets."

Common variants are present throughout the population, so researchers suspect they work in combination with environmental factors. "It may be that neuronal function is perfectly fine under most circumstances, but if a body of neurons gets stressed in a certain way that can tip a person over," says Sullivan.

To discern the environmental contribution to schizophrenia's genetic risk, the PGC is collaborating with epidemiologists Preben Bo Mortensen at Aarhus University in Denmark and John McGrath at the University of Queensland in Australia. Mortensen has access to genetic information from 2.2 million people born in Denmark since 1981 who had blood collected at birth. His group aims to develop a polygenic risk score for 85,000 of them, predicting each individual's likelihood of developing schizophrenia from their composition of common variants. The researchers will also survey other factors linked to schizophrenia risk, such as low birth weight or low vitamin D levels, enabling them to tease apart the genetic and environmental contributions to schizophrenia.

OPENING THE EXOME

Despite the explosion of knowledge about common variants, some researchers continue to see the most promise in rare mutations. These mutations can reveal pathways that are undeniably implicated in the disorder, says David Goldstein, a geneticist at Duke University in Durham, North Carolina. "The part of the genetic architecture I'm interested in is where a smaller set of higher-impact mutations imply risk that I can study biologically and use as guidance for drug development," he says.

To search for rare variants that remove, insert or change single DNA nucleotides, researchers are sequencing exomes, the protein-coding regions of the genome. Today's exome studies still have too few sequences to reveal a statistically significant role for a particular gene. But they do implicate certain pathways, such as the development and migration of new neurons and signalling through NMDA (*N*-methyl-D-aspartate) receptors. The studies also suggest an important role for proteins at the receiving ends of synapses (the postsynaptic density).

The two largest exome-sequencing studies in schizophrenia so far were published in *Nature* in January 2014. In one, researchers sequenced the exomes of 623 people with schizophrenia and their unaffected parents². The other looked at the exomes of 2,546 genes associated with schizophrenia or other psychiatric disorders in more than 5,000 people, divided between those with schizophrenia and controls³. Both

studies found several hits in genes regulated by the fragile X mental retardation protein, which is missing in fragile X syndrome, a disorder that can lead to intellectual disability and autism.

Rare spontaneous mutations may also take the form of copy number variants (CNVs) — duplications or deletions of stretches of a chromosome. Roughly one in 20 people with schizophrenia have CNVs of more than 100 kilobases⁴. Each CNV spans a number of genes, one or many of which may be associated with the disorder. Those implicated in schizophrenia tend to encompass genes that also play a role in the postsynaptic density and NMDA signalling.

The most well known of these CNVs, a 3-megabase deletion in the 22q11.2 chromosomal region, leads to schizophrenia or psy-

"Schizophrenia genetics — a reality at last!"

chosis about 25% of the time. One recent study suggests that duplication of this region may protect against schizophrenia while increasing the risk of autism⁵, demonstrating the complex and overlapping relationships in the genetics of psychiatric disorders.

Indeed, different genetic analyses are highlighting some of the same genes and pathways. Both the PGC GWAS and the recent exome studies, for example, identified multiple hits for genes involved in the L-type calcium channel, which controls muscle and cardiac activity. The multiple hits in these calcium-channel subunits, using different genetic analyses, suggest that these may be the most promising potential therapeutic targets discovered so far.

Researchers hope to find more targets as they identify further gene candidates. To push their statistical power even higher, a team at PGC designed the Psych Chip — a device that makes it cheaper to screen the genomes of a large number of individuals (roughly US\$50 per person). The chip covers 60,000 common and rare variants that were implicated in schizophrenia in previous studies but fell short of statistical significance. It also covers genes within CNVs. The researchers aim to test 100,000 individuals with psychiatric disorders in the next year. Based on results so far, this analysis should generate at least four more statistically significant hits for every 1,000 individuals, says Ripke.

Two years from now, researchers will need a larger T-shirt to flaunt their findings. "It's becoming clear that we are far from the end of the discovery step," says Ripke. "A hundred hits will not be the answer. There will be many hundreds if not thousands, which frightens many people. But we are definitely not slowing down." ■

Jessica Wright is a freelance science writer based in New York City.

1. Ripke, S. *et al. Nature Genet.* **45**, 1150–1159 (2013).
2. Fromer, M. *et al. Nature* **506**, 179–184 (2014).
3. Purcell, S. M. *et al. Nature* **506**, 185–190 (2014).
4. Kirov, G. *et al. Mol. Psychiatry* **17**, 142–153 (2012).
5. Rees, E. *et al. Mol. Psychiatry* **19**, 37–40 (2014).



Rodents are used to test drugs designed to correct cognitive deficits in people with schizophrenia.

DRUG DEVELOPMENT

The modelling challenge

Researchers have made good progress with animal tests for cognition. The next step is to devise a rodent model for drug development.

BY ALLA KATSNELSON

When Patricio O'Donnell started his lab in 1997 at Albany Medical College in New York, schizophrenia research seemed to be moving forward after a long period of stagnation.

For several decades, attention had focused on the idea that the disease was caused by elevated dopamine levels in the brain, particularly in the striatum, a nugget of brain tissue nestled under the cortex. But by the 1990s, the dopamine hypothesis was proving inadequate to fully explain the disease. *In vivo* imaging with computed tomography and magnetic resonance imaging, and data from post-mortem studies in people with schizophrenia, pointed to cortical effects and implicated other neurotransmitter systems, such as glutamate and

serotonin. Biologists were also learning to create transgenic mouse models of the disease, providing a set of tools with which to investigate genetic and aetiological factors.

This proliferation of ideas opened the doors to fresh avenues of investigation, as well as new ways of modelling the disease in animals. The classical rodent model for schizophrenia had involved administering amphetamine — which ramps up dopamine levels in synapses and can cause hallucinations and delusions in people — and then measuring behaviours such as hyperactivity or passive avoidance. Researchers then began using other pharmacological agents, such as phencyclidine (PCP) and ketamine, which interfere with the

glutamate receptor NMDA, or knocking out specific genes, such as those that code for neurotransmitter receptors.

O'Donnell was captivated by one particular approach that entails chemically damaging part of the hippocampus in newborn animals. Rodent pups who receive this intervention initially appear normal, but they later become socially withdrawn and overly responsive to stress, and their cortices show some of the physiological changes observed in post-mortem tissue. These findings reflect the now-accepted hypothesis that schizophrenia is a developmental disorder. "What made it attractive to me was that even though the lesion was done really early in development, all these behavioural abnormalities didn't emerge until adolescence," says O'Donnell, now vice-president and head of psychiatry and behavioural disorders at Pfizer Neuroscience in Cambridge, Massachusetts.

Attempts to mimic the symptoms or underlying physiological effects of schizophrenia have multiplied since then. As genome-wide association studies began pinpointing genes associated with the disease (see 'Unraveling complexity', page S6), researchers have knocked out or perturbed those genes in mice. Others have looked at environmental stressors, such as prenatal infections, early social isolation or stress, or lesions, like the ones O'Donnell used in the hypothalamus. And others give animals drugs such as amphetamine, ketamine and PCP. Some researchers have begun combining some of these manipulations — delivering an environmental stressor, for example, to mice lacking a particular schizophrenia-associated gene.

But when it comes to developing drugs, which are the best models? "There are so many risk factors, so many changes — the dopaminergic system, serotonin, GABA, you name it," says Thomas Steckler, a behavioural neuroscientist at Janssen Research and Development in Beerse, Belgium, who works in drug discovery. "We just don't know what best to model."

BEYOND 'ME TOO'

An animal model has two components: the manipulation that mimics some aspect of the disease, and the test — behavioural or otherwise — that measures the deficit that the manipulation produces. Schizophrenia researchers are getting a handle on the latter, but there's much less clarity regarding the former.

For years, schizophrenia drug discovery efforts were aimed at treatments for psychosis and delusions because new compounds could be tested against the two classes of antipsychotics — the only medicines licensed to treat the disease — that were already on the market. Unsurprisingly, that strategy was unproductive. "If you have models and all they are doing is comparing new approaches with what's already there, how can you move forward?" says neuroscientist Bitu Moghaddam of the University of Pittsburgh in Pennsylvania. "All

➔ NATURE.COM
More on using
animal models to
fight disease here:
go.nature.com/q3pbJc

we have done is make 'me too' drugs for the past 40 or 50 years."

But in the past 15 years, studies have shown that the disease has other core features that are at least as important as psychosis: difficulties with working memory, poor attention, and other cognitive symptoms. Drug discovery efforts have shifted accordingly, making some aspects of animal modelling more straightforward. "Modelling psychosis or negative symptoms [such as lack of emotion or motivation] is tremendously difficult, whereas we know an awful lot about cognition and brain systems in animals, and we know a lot of it translates to humans," says Trevor Robbins, a cognitive neuroscientist at the University of Cambridge, UK.

His lab and others have devised behavioural tests of cognitive processes in animals that can be matched to tests of cognitive processes in people with schizophrenia. One test, for example, probes attention by how accurately rodents respond to a brief stimulus in one of five locations in a cage, and how well they combat impulsivity by not responding before the stimulus occurs¹. Two major collaborative projects aimed at boosting drug discovery efforts for cognitive symptoms of the disease are working on validating these tasks so researchers can be sure they are using comparable procedures and tapping into the right cognitive modalities. In the United States, researchers working within an effort called CNTRICS (Cognitive Neuroscience Treatment Research to Improve Cognition in Schizophrenia), funded by the National Institute of Mental Health, recently agreed on a handful of tasks that most accurately reflect each of six different cognitive or behavioural domains, such as working memory, executive function and motivation².

Meanwhile, as part of a €20 million (US\$28 million), five-year public-private partnership in Europe launched in 2009, eight key industry players including Janssen, Eli Lilly and Novartis have joined forces to standardize a set of behavioural tasks that can be used in preclinical testing of new drugs. The collaboration, called NEWMEDS (Novel Methods leading to New Medications in Depression and Schizophrenia), had multiple participants do the same tasks to make sure their procedures aligned. "We started off with object recognition," says Mark Tricklebank, a behavioural neuroscientist at Eli Lilly in Surrey, UK, who is coordinating the animal model part of NEWMEDS. "We found that everybody was doing it in a slightly different way," he says. Researchers were using different-sized testing spaces and objects, for example, or testing at different times of the day. It took NEWMEDS several months per task to hammer out shared protocols. There is considerable overlap between the tests used by CNTRICS and NEWMEDS, but because

NEWMEDS' goal is industry testing, it focused on touchscreen versions of the tasks.

"We are at a stage where we can reasonably well test our animals in paradigms that we at least think tap into the same cognitive domains" as they do in humans, says Steckler. Solving the other side of the animal model equation, however — how to produce a deficit in an animal that a drug must reverse — will be much tougher, he says.

CAUSES AND DEFICITS

Having such a wide range of manipulations available should provide the means to test biologically based hypotheses, says O'Donnell. One hallmark of the disease, for example, is the loss of a specific population of interneurons in the frontal cortex, a feature that is also present in many animal models. To determine whether targeting this mechanism will ameliorate disease symptoms, drug candidates need to be tested on this particular feature in animal models. But so far, he says, industry tests its drugs largely with older classical models, inducing symptoms with pharmacological agents. "Even though companies are adopting ideas coming from animal work," O'Donnell says, "the use of biology-specific models is not there yet."

That's changing, says Steven Siegel, director of the translational neuroscience programme at the University of Pennsylvania in Philadelphia. He says that industry researchers have reached out to his group and others to discuss potentially useful and reproducible models, and some companies are investing heavily in setting them up.

On the flip side, however, industry researchers contend that studies from academia are not always reproducible. Steckler cites the case

where a university scientist identifies a cognitive deficit in a mouse in which a schizophrenia risk gene was knocked out. "That's your deliverable in academia; what you do with these data is you publish them," Steckler says. In industry, by contrast, that result must be more repeatable and robust, he says. But there's a catch: with no compound available that improves cognition in people with schizophrenia, it's unclear what to test models against.

In addition to its work on cognitive paradigms, NEWMEDS is trying to tackle the animal models issue from this end. The group has focused on trying to standardize a lesion manipulation in which pregnant rats or mice are injected with methylazoxymethanol acetate, which briefly blocks cell division in the developing embryos.

"The difficulty is diagnosing which animals to include so you have a deficit to correct in the drug trial."

As with hippocampal lesions, in adolescence the animals begin to show many hallmarks of schizophrenia, such as cortical thinning, ventricular enlargement, the loss of a subset of interneurons

in the frontal cortex, and cognitive changes.

Despite using animals that were bred and injected at the same facility maintaining protocols that are as close as possible among participating companies, the outcome varies — and some animals don't show the expected behavioural changes at all. That creates a problem when it's time to test them a year later. "The difficulty is then accurately diagnosing which animals to include in the drug treatment, so that you can be sure to have a deficit to correct in the pharmacological trial," Tricklebank says. The group is incorporating brain activation and other physiological measures that could also reflect whether a drug is having an effect. Results with genetic models have faced similar challenges, Tricklebank says, but he hopes the group will agree some standardized manipulations before its funding ends later in 2014.

Even if all these issues were resolved, says Tricklebank, there's still a big leap as to whether drug activity in an animal model predicts its activity in humans. Still, validating every step offers the best hope for success. Robbins, the academic leader of NEWMEDS' animal model efforts, agrees. "As far as I'm concerned, we're just at the beginning of a real scientific attack on these issues." ■

Alla Katsnelson is a freelance science writer based in Northampton, Massachusetts.



Monitoring the behavioural patterns of humans (top) and mice.

1. Amitai, N. & Markou, A. *Biol. Psychiatry* **68**, 5–16 (2010).
2. Moore, H. et al. *Neurosci. Biobehav. Rev.* **37**, 2087–2091 (2013).



Daniel Horan was miserable until magnetic brain stimulation helped him get his life back on track.

THERAPEUTICS

Negative feedback

Schizophrenia debilitates not just by psychosis but by depriving people of the ability to feel pleasure.

BY ELIE DOLGIN

Daniel Horan was 20 years old and studying to be a chemical engineer at the University of New Hampshire in Durham when he started to believe he had the power to read minds. He was also hearing voices and had an elevated sense of self-importance. After a confrontation with his roommate in 1998 — around the time that Horan believed he was “the devil in hell surrounded by fallen angels” — Horan was suspended from the university. Psychiatrists soon diagnosed him with schizophrenia.

Horan started taking an antipsychotic drug called risperidone, which made him go from feeling “sky high” to “extremely depressed”, he recalls. Blaming the emotional downturn on the drugs, he stopped taking the risperidone. His delusions soon returned and he was hospitalized again. Horan switched medications to another antipsychotic called olanzapine.

Once more, he felt a lack of motivation and even suicidal thoughts. He stopped taking the olanzapine but soon found himself at an inpatient psychiatric centre.

Throughout these episodes in and out of the hospital, Horan assumed that his blunted emotions and lack of motivation were simply side effects of the antipsychotic drug therapy. He was wrong. Such ‘negative’ symptoms are actually an integral and primary feature of the disease, experienced by around 20% of people with schizophrenia. “Everyone talks about the positive symptoms — the hallucinations, the delusions and the paranoia,” says Horan, now 36 and an auctioneer selling antique clocks and watches in Windham, New Hampshire. “I didn’t realize there was another side to schizophrenia.”

Flat affect, social withdrawal, inability to feel pleasure — these are often a huge barrier preventing people with schizophrenia from living independently, holding down jobs,

establishing personal relationships or managing everyday social situations. But there are no existing therapies, pharmaceutical or otherwise, that are approved to treat the negative symptoms of schizophrenia.

“We have very little that we’re able to offer patients for these symptoms,” admits Steven Zalcman, head of the Clinical Neuroscience Research Branch at the US National Institute of Mental Health in Rockville, Maryland. Antidepressants are better than nothing, but the benefit they provide is usually modest at best.

Fortunately, new therapeutic options could be on the way. Horan, for his part, recently participated in a clinical trial designed to test whether a non-invasive magnetic brain-stimulation technique could help with the negative symptoms of his disease. Others are involved in studies evaluating new drugs or innovative forms of psychosocial treatment.

Success with one of more of these therapies in mitigating the recalcitrant negative symptoms of schizophrenia should dramatically enhance the total management of the disease, says Stephen Marder, director of the Section on Psychosis at the University of California, Los Angeles, Neuropsychiatric Institute. “It would transform treatment.”

NO GOOD

The arrival of a new class of ‘atypical’ antipsychotics in the late 1980s was supposed to solve the problem of negative symptoms. But as Horan’s experience shows, these drugs were not the solution. Indeed, despite early reports of benefit, subsequent meta-analyses and well-controlled trials showed no reduction in primary negative symptoms from these agents.

Over the years, many other types of medication have been examined as potential adjunct treatments for negative symptoms. Nearly all have failed. The most recent drug flops came from Targacept in Winston-Salem, North Carolina, and from the Swiss pharmaceutical giant Roche. Each had experimental compounds in mid- or late-stage development for what many considered to be the two most promising drug targets for controlling negative symptoms. In recent months, each released trial data showing that their drugs proved no better than placebo pills.

Results like these “make the task of finding truly novel treatments for negative symptoms of schizophrenia that much more daunting”, says Jeffrey Lieberman, psychiatrist in chief at New York Presbyterian Hospital–Columbia University Medical Center in New York. Yet even though the specific drugs may have had their flaws, Lieberman doesn’t think the targets are necessarily the wrong ones. “We must persevere,” he says.

The target pursued by Targacept was the alpha-7 nicotinic acetylcholine receptor, which is activated by the neurotransmitter acetylcholine and by nicotine. Studies show that individuals with schizophrenia have serious

reductions in alpha-7 receptors in several regions of the brain. Many people with schizophrenia overcome their lack of alpha-7 receptors by smoking cigarettes — often heavily.

Research has found that some schizophrenia patients smoke to self-medicate, as it relieves their negative symptoms. Nicotine is hardly an ideal therapeutic agent, however. As well as the health risks of smoking, nicotine is an addictive substance to which people develop tolerance. This limits the long-term benefits of nicotine and it's why scientists have been working on less toxic and more chronically effective treatments that make alpha-7 receptors more responsive to naturally occurring acetylcholine.

The proof-of-principle that activating the alpha-7 receptor with a small-molecule drug could ameliorate negative symptoms came from Robert Freedman, a psychiatrist at the University of Colorado School of Medicine in Aurora. Freedman and his colleagues ran a trial that involved 31 people with schizophrenia who received either placebo pills or twice-a-day doses of DMXB-A, which is a partial agonist of the alpha-7 receptor¹. The researchers expected the treatment to improve cognitive impairments associated with schizophrenia, such as trouble focusing or problems with working memory, but they observed more consistent benefits in negative symptoms.

Targacept initially saw the same effect with its activator of the alpha-7 receptor. In a 12-week, 185-person trial led by Lieberman, treatment with the Targacept drug led to significant improvements in negative symptoms compared with placebo². But in a 477-person follow-up study, the drug offered no such benefit. Targacept officials announced the failed trial results in December 2013.

TRICKY BUSINESS

Despite the clinical setback, Lieberman still sees potential in activating the alpha-7 receptor. "It's a viable target," he says, "but it's very tricky to engineer a compound that will be optimally effective." Freedman agrees. He thinks Targacept's once-a-day drug was probably too strong and long acting. Thus it engaged the alpha-7 receptor but did not disassociate fast enough, leading to a diminished response known as desensitization.

Although drug companies generally prefer to make once-a-day pills, patients in this setting might need to take shorter-acting drugs more often. "Maybe this therapeutic principle will work best if you get a couple of good hits of drug a day," says Freedman, who is now running an 80-person trial testing DMXB-A given four times per day. A handful of other companies have alpha-7 activators in various stages of clinical development, but they are all now being pursued primarily for cognitive enhancement, rather than to alleviate negative symptoms, says Harry Tracy, the editor and publisher of *NeuroPerspective*, a monthly

publication focused on treatments for neurological and psychiatric diseases.

Roche, meanwhile, was banking on a different therapeutic target: the glycine transporter 1 (GlyT1) protein found on glial cells in the brain. When the protein is blocked, more of the amino acid glycine stays in the synapse. The glycine then activates NMDA (*N*-methyl-D-aspartate) receptors, which are involved in the signalling of glutamate, an excitatory neurotransmitter that's deficient in people with schizophrenia and is thought to underlie many of the negative and other symptoms of the disease.

In 2010, Roche presented data from a 323-person, phase II trial demonstrating improvements in the negative symptoms of participants who took the company's experimental GlyT1 inhibitor bitopertin for eight weeks rather than a placebo pill³. Decision Resources Group, a market-research firm in Burlington, Massachusetts, started to forecast blockbuster sales for the agent. But in January 2014, Roche dropped a bombshell. The company released top-line results for two of the three pivotal follow-up trials designed to test bitopertin's impact on negative symptoms. In both phase III studies, adding bitopertin to antipsychotic therapy for 24 weeks did not significantly reduce negative symptoms compared with placebo.

As *Nature* went to press, Roche was awaiting data from its remaining studies — one focused on predominant negative symptoms, three others on suboptimally controlled positive symptoms — before deciding the fate of bitopertin. A few other drug companies are also pursuing different ways of modulating glutamate signalling to combat negative symptoms.

NOW WE'RE TALKING

With pharmacological efforts to treat negative symptoms facing so many stumbling blocks, a number of researchers have begun to test non-drug interventions. These include forms of psychological techniques used more commonly by mental-health professionals in other disease settings, such as cognitive behavioural therapy and social skills training, as well as less traditional methods, including art therapy and so-called loving-kindness meditation.

Ann Kring and her colleagues at the University of California, Berkeley, recently surveyed the literature on psychosocial therapies for negative symptoms⁴. Although empirical data are limited, the researchers concluded that these treatment techniques hold promise as add-ons to medications. "Negative symptoms are an area where psychosocial treatments could be a wonderful adjunct," Kring says.

So could brain stimulation. In December 2013, as part of a clinical trial at the Beth Israel Deaconess Medical Center in Boston, Massachusetts, Horan received ten treatments over five days of transcranial magnetic stimulation. This brain-activating therapy involves placing an electromagnetic coil above the scalp and

releasing magnetic pulses that cause nerve cells to fire. By altering the wiring of dysfunctional neural networks, this treatment, when administered over the left prefrontal cortex to people with schizophrenia, has been shown to moderately alleviate negative symptoms⁵. Now, a team led by neurologist Alvaro Pascual-Leone at Beth Israel Deaconess is testing

"One woman started living independently, holding down a job, making friends — all things she didn't do before."

whether such stimulation works even better when directed at the cerebellum, a brain region involved with regulating emotion and its display.

In a preliminary, eight-person trial⁶, the therapy proved "revolutionary" for

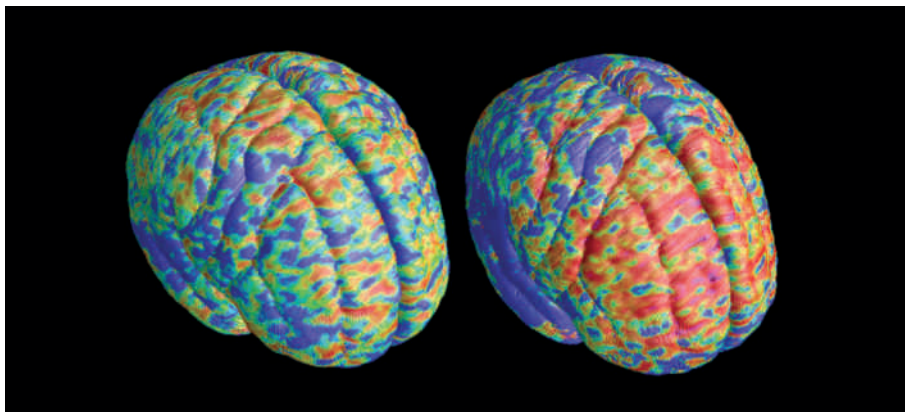
some participants, says Mark Halko, one of Pascual-Leone's colleagues. One woman, he says, "went from living at home to living independently, holding down a job, making friends — all things she didn't do before the trial." In the follow-up trial, which involves a sham treatment to test for a placebo effect, similar dramatic responses can be seen in participants like Horan. After the first few treatments with magnetic stimulation, "it was like a miracle," says Horan, who learned after-the-fact that he received the real treatment during the trial. "I had verve, pep, enthusiasm."

The treatment wasn't a cure, though. In the months after the trial, the beneficial effects gradually started to wear off for Horan. The technique is approved in the United States only as a treatment for major depression or migraine headaches. But fortunately for Horan, his doctors agreed to continue administering the therapy off-label. In March 2014, Horan received six more sessions of brain stimulation over three days at Beth Israel Deaconess. Although "less intense" than his first experience, Horan says, "there was definitely an improvement."

After years of working part time and receiving disability benefits from the government, Horan can now see a future where he's self-sufficient. He's even trying to buy out his business partner at the auction house. The magnetic stimulation "gives me a little glimmer of hope that I might at least have the health I need," Horan says. "It's given me a new lease on life." ■

Elie Dolgin is senior news editor at *Nature Medicine* in Cambridge, Massachusetts.

1. Freedman, R. *et al.* *Am. J. Psychiatry* **165**, 1040–1047 (2008).
2. Lieberman, J. A. *et al.* *Neuropsychopharmacology* **38**, 968–975 (2013).
3. Umbricht, D. *et al.* *Neuropsychopharmacology* **35**, s320–s321 (2010).
4. Elis, O., Caponigro, J. M. & Kring, A. M. *Clin. Psychol. Rev.* **33**, 914–928 (2013).
5. Prikryl, R. & Kucerova, H. P. *J. ECT* **29**, 67–74 (2013).
6. Demirtas-Tatlidede, A. *et al.* *Schizophr. Res.* **124**, 91–100 (2010).



The brain of someone developing schizophrenia (right) typically shrinks more rapidly than normal (red colours indicate the highest rates of contraction).

PREVENTION

Before the break

Paying attention to risk factors and warning signs could avert some cases of schizophrenia — or at least better prepare people for what's to come.

BY MICHELE SOLIS

Like any mother, Farrell Adrian worried about her child, Matt. He started off as a happy, sweet boy, but grew unusually sad at the age of five, when his father left the family. He developed problems with thinking and focus at school, and as a teenager he became depressed and addicted to drugs. Along the way, Farrell took him to see psychologists and other specialists, hoping that if they could solve what she thought was an emotional issue about his father, the old Matt would return.

When Matt was 17, he was finally diagnosed with schizophrenia. Twenty years on — after a bumpy road of hospitalization, therapy and antipsychotic medicines — Matt lives on his own in Seattle, Washington, works part-time at a football stadium, and is close to his family. Even though it took more than a decade to figure out what was wrong, Farrell says she feels lucky. “We got help sooner than most people do,” she says.

Some researchers are now proposing that others could be helped much sooner, by being alert to signs that unfold during the months or even years preceding psychosis. A growing number of studies suggest that treating these early, muted signs can halve the number of people who later develop a break from reality. Researchers are also hunting for more concrete signs of risk in the brain, blood and saliva to better identify those at risk. Pre-emptive treatment could delay, or even avert, schizophrenia,

and bring psychiatry into the realm of preventive medicine.

“Traditionally psychiatry offers too little too late,” says Patrick McGorry, a psychiatrist at the University of Melbourne in Australia who advocates earlier intervention. “But now we see that the path to psychosis is a lot more sensitive to intervention than we used to think.”

If better predictors of risk are found, and the successes of small treatment trials can be replicated on a larger scale, pre-emptive treatment could become an accepted practice in psychiatry, says Jeffrey Lieberman, a psychiatrist at Columbia University in New York and president of the American Psychiatric Association. “Then it could be a game changer and reduce the burden on people with schizophrenia, their families and society.”

SEEDS OF PSYCHOSIS

Evading schizophrenia is not a new idea. Epidemiological studies have identified environmental risk factors that may contribute to errors in early brain development that are thought to underlie schizophrenia. For example, improving nutrition and avoiding infection during pregnancy may prevent some cases of the illness, and researchers are considering the possibility that getting the correct amount of vitamin D in early life could decrease the risk of schizophrenia (see

‘Born at risk’). Even when schizophrenia has taken hold, early treatment after the first episode of psychosis can limit the severity of the illness and increase the chances of recovery.

McGorry and others have taken the idea of early treatment further, proposing to intervene at the first suggestive signs of psychosis. A person might become suspicious, or start to hear a voice, but still recognize that these things are not tethered to reality. They may withdraw from friends and family, or have trouble focusing at school or work. The criteria for this ‘at risk’ category differ slightly between researchers, but they all require that a person is sufficiently distressed by their symptoms to seek help.

About one-third of people in this at-risk category develop psychosis within three years, and most are diagnosed with schizophrenia. A version of this at-risk category, called attenuated psychosis syndrome (APS), was considered for inclusion as a new diagnosis in the recent fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)*, one of the most widely used inventories of mental illnesses. But most people with APS do not actually develop full-blown psychosis, so after much discussion the syndrome was not included in *DSM-5*.

Despite these diagnostic uncertainties, people deemed at risk for psychosis based on their behaviour carry marks of their vulnerability in their brains. For example, an excess of the chemical messenger dopamine in a part of the brain called the striatum is thought to drive psychosis. A 2011 study¹ found that people at risk for psychosis had a surplus of dopamine in the striatum compared with healthy controls, and those with the highest levels were most likely to develop psychosis in the next three years. Other studies have found that the prefrontal cortex — the brain’s supervisory centre — shrinks more than usual in people at risk during typical ageing, although not as much as in people with schizophrenia.

“We think we’re detecting the early, leading-edge indicators of the same processes at work in schizophrenia,” says Tyrone Cannon, a neuroscientist at Yale University in New Haven, Connecticut, who conducted the studies on the prefrontal cortex.

TREATMENTS ON TRIAL

So far, 11 randomized, controlled trials have found that treating people at risk results in fewer people developing psychosis within one year. Mostly small, the trials tested interventions such as low doses of antipsychotics, cognitive behavioural therapy (CBT) and omega-3 fatty acid supplements.

The earliest trials found that antipsychotics had beneficial effects, but these are no longer seen as a first line of treatment for people at risk. “There’s a lot of sentiment, I think wisely so, that antipsychotics can easily do more harm than good for people who are

➔ **NATURE.COM**
Find out more
about attenuated
psychosis syndrome:
go.nature.com/2HJP3S

not actually on the verge of developing a full psychotic illness,” says William Carpenter, a psychiatrist at the University of Maryland in Baltimore. Carpenter oversaw the discussion of whether to include APS in *DSM-5*.

Instead, studies in the past five years have emphasized safer treatments such as CBT, an approach that teaches people to recognize their own patterns of thinking and to reappraise situations. For example, someone with APS who jumps to conclusions about the bad intentions of others will learn from CBT to recognize this, and not to trust their first suspicions but find alternative explanations for someone's behaviour. In 2012, the largest CBT trial to date confirmed its ability to stave off psychosis in 201 people at risk: 12% of the group receiving CBT developed psychosis within 18 months, compared with 24% in the group that did not get CBT.

A more peculiar, and preliminary, finding points to beneficial effects of omega-3 fatty acid supplements in the form of fish oil. Omega-3 fatty acids are components of neuronal membranes that also dampen inflammation and oxidative stress — both suspected of contributing to schizophrenia. A 2010 study² led by Paul Amminger, a psychiatrist working at the Medical University of Vienna in Austria and with McGorry in Melbourne, reported that only 5% of people at risk who took four fish oil capsules a day for 12 weeks developed psychosis, compared with 28% who took a placebo. Unpublished results suggest that this regimen staves off psychosis for at least six years, and a larger trial is now seeking to replicate the finding.

“People always ask me if omega-3 fatty acids will be as good as antipsychotics, but it seems they're even better,” Amminger says. Unlike the early antipsychotic trials, he notes, those taking the supplements also showed improvements in the way they functioned in daily life.

“If they replicate that study, it may be the most important thing since Thorazine,” Carpenter says, referring to chlorpromazine, the first antipsychotic drug, which was introduced in the 1950s.

RECOGNIZING RISK

To get a better idea of who could benefit from treatment, researchers are looking for other

risk indicators that could more precisely flag someone on the path to psychosis. One comes from a survey of cognitive abilities, which accurately predicted 91% of people who developed psychosis and 89% of those who did not³. Brain imaging and electroencephalography also reveal patterns that start to differentiate between those who later develop psychosis and those who do not.

Beyond the brain, levels of the stress hormone cortisol and profiles of gene expression and protein abundance in blood cells are also likely candidates. For example, research in 2013 from the North American Prodrome Longitudinal Study (NAPLS), a consortium studying the early signs of psychosis and its prevention, found higher cortisol levels in

“People always ask me if omega-3 fatty acids will be as good as antipsychotics, but it seems they're even better.”

people at risk of psychosis⁴ — especially those who later developed psychosis. In the future, someone with APS could be evaluated by using a computerized cognitive test, a saliva swab and a blood test. “Based on these multiple lines of information, an algorithm would spit out a

number that would provide enough predictive power for a clinician to help that person,” says Cannon, who directs NAPLS.

Until then, increasing awareness among healthcare professionals of the early behavioural signs of psychosis risk could help a lot. Given that most mental illnesses appear between 12 and 25 years of age, McGorry has spearheaded the creation of youth mental-health centres in Australia that provide a friendly place for young people to get support for any kind of problem, not just the early stages of psychosis.

Even in the United States, where people typically have to reach a crisis point before receiving mental health care, things may be changing, thanks in part to the horror of recent gun violence carried out by people presumed to be mentally ill. Prompted by the tragic shootings in Newtown, Connecticut, in 2012, the state of Maryland has recently funded the creation of a youth mental-health programme similar to those in Australia. “I expect a mushrooming of these sorts of early identification and intervention centres,” Carpenter says, “which won't wait on the science.” ■

Michele Solis is a freelance science writer based in Seattle, Washington.

1. Howes, O. D. et al. *Am. J. Psychiatry* **168**, 1311–1317 (2011).
2. Amminger, G. P. et al. *Arch. Gen. Psychiatry* **67**, 146–154 (2010).
3. Koutsouleris, N. et al. *Schizophr. Bull.* **38**, 1200–1215 (2012).
4. Walker, E. F. et al. *Biol. Psychiatry* **74**, 410–417 (2013).

BORN AT RISK

Prevention in the womb

Schizophrenia strikes nearly 1% of the world's population, but affects certain people more than others. Someone born during the winter months, for example, has a higher risk of schizophrenia than someone born in the summer. And dark-skinned people living in northern countries are three to four times more likely to develop schizophrenia than light-skinned natives. “Those clues scream out that there must be modifiable risk factors underpinning those gradients, and we need to find out what they are,” says John McGrath, an epidemiologist at the University of Queensland in Brisbane, Australia.

McGrath has been investigating the case of dark-skinned migrants, starting with the hypothesis that their increased risk may be due to a lack of vitamin D, which is important for proper brain development and is often lacking in dark-skinned people in northern countries. He found that in Denmark, abnormal levels of vitamin D at birth were associated with an increased incidence of schizophrenia. If a larger study currently underway confirms this connection, McGrath thinks it might be worth monitoring vitamin D, particularly in dark-skinned migrant populations.

Seasonal infections such as influenza could partly explain the higher risks faced by those born in winter. Pregnant women may pass these infectious agents to their unborn child, where they could affect brain development. Alan Brown, a psychiatrist at Columbia University in New York, has found that while mothers were pregnant with babies who eventually developed schizophrenia, they showed abnormally high levels of antibodies to the influenza virus in their blood.

These findings point to the importance of nutrition and avoiding infection during pregnancy, which are already prominent messages for having a healthy baby in the developed world. These guidelines may already be reducing schizophrenia incidence in developed nations, and extending them to developing countries may bring further decreases.

“Many of the big medical advances — say, in cardiovascular disease or cancer — have come through changes in our diet and behaviour,” McGrath says. “There's no reason to think that mental illness would be any different.” **M.S.**



Fish oil capsules containing omega-3 fatty acids may be able to delay the onset of psychosis.



A rural lifestyle and lack of stigma might help Indians recover faster from schizophrenia.

DEVELOPING COUNTRIES

The outcomes paradox

Schizophrenia patients in developing countries seem to fare better than their Western counterparts. Researchers are keen to find out why.

BY T. V. PADMA

The pattern for most diseases is clear: the richer and more developed the country, the better the patient outcome. Schizophrenia appears to be different.

This paradox first came to light 40 years ago. Studies from Mauritius and Sri Lanka appeared to show better outcomes than developed countries: patients experienced fewer delusions and hallucinations, less disorganized speech, and improved social functioning. But these studies lacked standardized diagnostic criteria and assessment methods, and had varying attrition rates.

In the late 1960s, in an effort to standardize research methods, the World Health Organization (WHO) launched the first of the following three landmark international studies: the

International Pilot Study of Schizophrenia (IPSS); the Determinants of Outcomes of Severe Mental Disorders (DOSMeD); and the International Study of Schizophrenia (ISoS).

The IPSS enrolled a total of 1,202 patients in nine countries: three developing countries (Colombia, India and Nigeria) and six developed ones (Denmark, Taiwan, the United Kingdom, the United States, the Soviet Union and Czechoslovakia). The patients' outcomes were assessed by using three indicators — the percentage of time with psychosis symptoms, the type of remission after each episode, and the degree of social impairment — and were classified on a scale of one (best) to seven (worst). At the five-year follow-up,

India had the most success, with 42% of cases reporting 'best' outcomes, followed by Nigeria with 33% of cases. By contrast, the developed countries had poor showings: 'best' outcomes were seen in only 17% of cases in the United States, and in fewer than 10% in the other developed nations.

Beginning in the early 1980s, the DOSMeD study examined schizophrenia incidence, prevalence and outcomes in 12 centres across 10 countries (the IPSS countries plus Ireland). Its 1,379 patients were assigned to 1 of 9 categories, ranging from a single episode of psychotic illness followed by complete remission, to continuing illness. The study found that developing countries had higher rates of complete recovery: an average of 37% compared with 15.5% in developed countries. The rates of chronic illness, however, were similar: 11.1% in developing and 17.4% in developed countries. Patients in developing countries experienced longer periods of unimpaired social functioning, even though far fewer of them were on continuous antipsychotic medication. The researchers concluded that "a strong case can be made for a real pervasive influence of a powerful factor which can be referred to as 'culture' in the context of gene–environment interactions that influence a disease. The contribution of the present study is not in providing the answer but in clearly demonstrating the existence of the question."

The ISoS study returned to the IPSS and DOSMeD patients after 15 and 25 years, and included two other groups, to test whether the better outcomes observed in the previous studies continued in the long term. It traced about 75% of the patients, finding that about half had favourable outcomes, but there was wide variation across different centres. The study concluded that socio-cultural conditions can modify the long-term disease course, and that early intervention programmes with social and pharmacological treatments could have longer-term benefits.

VARIABLE RESULTS

These WHO studies were far from perfect, however. The IPSS studies looked only at patients in psychiatric facilities, and so may not represent the wider population. The DOSMeD study, in contrast, actively sought out patients not in such facilities. All the studies experienced high attrition rates from participants in developing countries, for reasons ranging from premature death to losing track of patients. Unfortunately, these drop-outs biased the results, as it was generally the 'worst' patients who dropped out. Patient selection and assessment of social functioning also varied. But despite these reservations, the idea that culture affects schizophrenia outcomes — and that developing countries perform better because of their culture — persisted.

In 2009, psychiatrist Parmanand Kulhara of the Postgraduate Institute of Medical

➤ NATURE.COM

For further research relating to schizophrenia recovery: go.nature.com/TnbiPm

Education and Research in Chandigarh, India, reviewed 58 schizophrenia studies to compare outcomes from developed and developing nations¹. With a few exceptions, he found that developing countries have a larger proportion of patients (50–60%) with good outcomes than developed countries in follow-up examinations after two and five years. These differences persisted beyond 15 years, but to a lesser extent. “Despite the controversies surrounding the favourable outcome hypothesis, we believe that the evidence arising from various studies cannot be simply dismissed,” his review concluded.

Nevertheless, some scientists remain sceptical. One of these is Vikram Patel, a public-health psychiatrist at the London School of Hygiene and Tropical Medicine. To look for trends within and between countries, Patel and his colleagues analysed² data from 23 longitudinal schizophrenia studies covering 11 low- and middle-income countries. They found inconsistent results based on different time frames that make direct comparisons difficult. For example, the proportion of people with chronic schizophrenia ranged from 4.5% over five years in India to 51.7% over 12 years in China. The disease pattern varied over time with patients shifting between best, worst and intermediate categories several times over the years. Disability and social outcomes also varied, being good in India and Indonesia, but poor in Brazil, China and Ethiopia. Patients in India fared better than in those countries, but good outcomes could not be generalized to other developing countries. Patel’s analysis concluded that it was “time to re-examine presumed wisdom about schizophrenia outcomes in low- and middle-income countries”.

CULTURAL CONTEXT

Even if the schizophrenia paradox is confined to India, the explanation is no clearer. “Patients seem to be doing better in poorer countries, despite limited resources such as health facilities and health infrastructure, and treatment facilities,” Kulhara says. This could be due to a different socio-cultural milieu in developing countries with a greater dependence on family members for care and support, and to better social support and social networking. However, Kulhara adds, “these variables have not received the research attention that they deserve.”

Other scientists are also trying to figure out the nature and role of cultural factors. “In the past three decades, the field has embraced the notion of ‘cultural differences’ as the reason for better outcomes in developing countries,” says Naren Rao, a neuroscientist at the Indian Institute of Science in Bangalore, India. “However, the nature of the cultural factors is not exactly known and culture remains a black box.” Even Kulhara, who has concluded that the difference in outcomes is real, notes in his

review that “culture should not be used as a synonym for unexplained variance”.

In trying to pin down these cultural factors, Rao points to studies showing that schizophrenia patients in India face fewer critical or hostile remarks than their equivalents in other countries — and says this may “contribute to the reason for better outcomes”. But an attentive family can cut both ways: in some cases, extended families provide additional care, but in others they prevent patients from seeking treatment for fear of stigma.

Some researchers attribute the improved outcomes to the less rigid nature of rural

“When it comes to a complex disorder like schizophrenia, it’s difficult to generalize for India, let alone for all developing countries.”

life in developing countries. Although people with schizophrenia in remote locations have less access to treatment, they tend to enjoy better social inclusion. Patel speculates that the disorganized rural labour markets offer more opportunities, such as field work for people with disabilities, which helps them integrate and overcome their illness. But he cautions against drawing too sweeping a conclusion. “When it comes to a complex disorder like schizophrenia,” he says, “it is difficult to generalize even for just India, let alone for all developing countries.”

Scientists are also examining the potential role of the widespread practice of yoga in India in the good outcomes for those with schizophrenia. Several studies from India over the past decade³ have reported that yoga has beneficial effects on schizophrenia, both with regard to ‘positive’ (psychotic) symptoms and ‘negative’ symptoms (such as lack of motivation or enjoyment), for which there are few effective treatments (see ‘Negative feedback’, page S10). Researchers have previously reported that both physical exercise and psychotherapy (including ‘mindfulness’ techniques such as meditation) can ease the symptoms of schizophrenia. “Since yoga involves both exercise and mindfulness, it seemed reasonable that it could be helpful too,” says Holger Cramer, a psychologist at the University of Duisburg-Essen in Germany.

Even so, the power of yoga remains questionable. Cramer recently led a review⁴ of five trials, including two from India, on the role of yoga in schizophrenia. He found no evidence that yoga has any effect on schizophrenia symptoms. Any positive effects seen in individual trials “could easily be just a statistical artefact”, he says.

RE-EXAMINING ASSUMPTIONS

This string of negative results makes any explanation for the Indian schizophrenia paradox even more obscure. Some scientists

are therefore taking a step back, to examine factors that influence outcomes in a particular setting. For example, two crucial indicators of improved social outcomes are employment and marriage. Marriage rates are higher for people with schizophrenia in India than in the United Kingdom, says Patel. But he argues that this should be juxtaposed against the fact that marriage rates in India are higher in general, partly due to family pressures for arranged marriages. “There are other linked factors, such as the huge numbers of untreated persons, human-rights abuses and high death rates” that such analyses ignore, says Thara Rangaswamy, director of the Schizophrenia Research Foundation in Chennai, India, a WHO collaborating centre for mental-health research and training. Rangaswamy has worked with Patel’s team in multicentre studies.

One of the more ambitious attempts to find out whether there really are differences in outcomes between locations is the three-year Intrepid study, in which Patel is involved. Its pilot phase, which takes place in India, Nigeria, and Trinidad and Tobago, runs until the end of 2014.

There is little directly comparable research on psychoses in low- and middle-income countries, partly as a result of logistical and methodological challenges, explains Craig Morgan, a psychologist at King’s College London who works on the Intrepid project. The pilot phase of Intrepid, Morgan says, is one of the first systematic attempts to develop and test robust psychoses research methods in diverse settings. This will lay the groundwork for a second phase aimed at studying actual causes and outcomes in large samples in each of the sites.

But conducting such research in countries with limited resources poses several challenges: recruiting representative groups of people with psychoses and population-based comparison groups; training staff and ensuring reliable assessments; and enabling cross-cultural comparison of assessments. Finding out whether there are indeed differences in symptoms, risk factors and outcomes in diverse settings “is important because differences provide a window into understanding the disorders more fully,” says Morgan. Rather than draw comparisons, researchers now want to untangle the factors that drive the onset and outcomes of schizophrenia — and they can then discover whether the paradox persists. ■

T. V. Padma is a science writer based in New Delhi, India.

1. Kulhara, P., Shah, R. & Grover, S. *Asian J. Psychiatry* **2**, 55–62 (2009).
2. Cohen, A. et al. *Schizophr. Bull.* **34**, 229–244 (2008).
3. Duraiswamy, G. et al. *Acta Psychiatr. Scand.* **116**, 226–232 (2007).
4. Cramer, H. et al. *BMC Psychiatry* **13**, 32 (2013).



AGEING

Live faster, die younger

People with schizophrenia show signs of accelerated ageing — a phenomenon that could lead researchers to a deeper understanding of the disease.

BY EMILY ANTHES

For decades, evidence has suggested that people with schizophrenia have shortened lifespans. In the early twentieth century, doctors observed that these patients tended to die younger and seemed less healthy than other patients in the same psychiatric hospitals. And today, schizophrenia patients appear to suffer from heart, lung and metabolic

problems at a disproportionate rate — and at startlingly young ages.

There are plenty of potential culprits: suicide, for example, along with the negative side effects of antipsychotic medications, substance abuse, smoking and poor health care — any one of which could explain why people with schizophrenia die 15–25 years earlier, on average, than those in the general population.

But some researchers believe that these

external factors do not fully account for the reduced longevity. They point out that the increased mortality among schizophrenia patients predates the widespread use of antipsychotics, encompasses diverse causes of death, and has been documented even among patients who are receiving high-quality medical care. These observations have led several investigators to propose a new explanation for the early deaths of those with schizophrenia: the ageing process itself speeds up. “Our hypothesis is that one of the reasons that they die younger is because they get old faster than the general population,” says Brian Kirkpatrick, a psychiatrist at the University of Nevada School of Medicine in Reno.

Accelerated ageing, Kirkpatrick and several other scientists say, may be a fundamental part of the schizophrenia pathology. This belief is supported by numerous studies of physiological, neurological and cognitive abnormalities in people with the disorder. Researchers are now trying to determine precisely how schizophrenia and ageing are related. They hope that fresh insights into the disease will help people with schizophrenia stay healthy longer, and possibly even point the way to better treatments.

MARKING TIME

As there’s no definitive test or biomarker for ageing, proponents of the accelerated-ageing hypothesis have been gathering support for their idea by piecing together a diverse array of research findings.

For example, scientists and doctors have noticed that several ailments associated with ageing, including type 2 diabetes and cardiovascular disease, are more common among people with schizophrenia (see ‘Schizophrenia’s health burden’), and appear earlier in life, than in the general population. The trouble is that antipsychotics commonly prescribed for schizophrenia can lead to weight gain and increase the risk of diabetes. So to control for these effects, studies need to focus on patients who have recently been diagnosed with schizophrenia and have not yet started taking medication.

Several studies of these antipsychotic-naïve patients have found that impaired glucose tolerance — a condition sometimes called prediabetes that is a risk factor for heart disease — is more common among patients with schizophrenia than in the general population¹. These results suggest that schizophrenia is associated with abnormal glucose metabolism, and that it is not solely a side effect of medication.

Kirkpatrick has also examined antipsychotic-naïve patients for other signs of premature ageing. In 2009, his team showed that these patients have an increased pulse pressure, a calculation of the difference between systolic and diastolic blood pressure². In general, pulse pressure increases with age and is a predictor of heart disease. And in 2011, Kirkpatrick’s team showed that antipsychotic-naïve men with schizophrenia have lower levels of testosterone,

a hormone that declines gradually with age, than age-matched controls³.

There are tantalizing signs in the genes, too. Telomeres, the DNA tails attached to the end of chromosomes that shorten each time a cell divides, are often used as a marker of biological ageing. Several small studies have documented shorter telomeres in patients with schizophrenia, but a larger 2013 study failed to replicate these results — in fact they found longer telomeres in people with the disease⁴. These apparently contradictory results are confounded by the fact the studies used different methodologies and examined different populations, ranging from antipsychotic-naïve patients to those who have had poor responses to the drugs.

Cognitive studies provide firmer support for the idea of accelerated ageing. Patients with schizophrenia seem to exhibit a premature deterioration in cognitive function, says Philip Harvey, a psychologist at the University of Miami Miller School of Medicine in Florida. Working with Christopher Bowie, a psychologist at Queen's University in Kingston, Ontario, Canada, Harvey tested processing speed and episodic memory — two areas of cognitive functioning that tend to worsen with age — and found that people with schizophrenia perform like subjects who are decades older⁵.

The reason for this decline could be physiological. The brain contains grey matter, comprising neurons, and white matter, made up of axons — the cellular projections that transmit signals between neurons. “As we age, we tend to lose our white-matter integrity,” says Peter Kochunov, a psychologist at the Maryland

Psychiatric Research Center at the University of Maryland in Catonsville. “The reason the brain slows down with age is because of this loss.”

Kochunov and his colleagues have examined white matter using diffusion tensor imaging, which measures the movement of water in the brain. (Water molecules tend to travel down the length of an axon.) They found that people with schizophrenia lose white-matter integrity at nearly twice the rate of controls⁶. “This is kind of alarming because in normal ageing, the loss of integrity is linked very strongly with basic cognitive domains such as processing speed,” he says. And the white-matter loss is specific to schizophrenia: Kochunov’s team did not find such rapid deterioration in people with major depression, another diagnosis associated with increased mortality and an elevated risk of a variety of ageing-related disorders.

FIVE-YEAR PLAN

The evidence so far suggests a correlation between schizophrenia and ageing, but many of these studies were small and based on single observations of patients and controls. Confirming the findings will mean following larger groups of subjects for many years, says Dilip Jeste, a psychiatrist and neurologist at the University of California, San Diego.

In 2012, Jeste began a five-year longitudinal study of accelerated ageing, enrolling more than 250 adults from 26 to 65 years of age, including individuals with and without schizophrenia. Each year, the researchers administer a battery of medical tests. For instance, they look for signs of inflammation and oxidative stress, which occurs when highly reactive, potentially damaging oxygen compounds flood the body’s cells. Both inflammation and oxidative stress tend to increase with age and are associated with a variety of chronic diseases. They have also been linked

to neurodegeneration and proposed as factors that might contribute to schizophrenia.

In addition, Jeste’s team will not only measure telomeres but also record their rate of shortening over time. “If schizophrenia is associated with accelerated ageing, you would expect that the shortening of telomeres over the five years in schizophrenia would be faster than that in healthy comparison subjects,” he says.

The scientists will also collect data on potentially confounding variables, including medications, substance use, smoking history and diet. They hope the results will show whether widespread accelerated ageing really does occur in patients with schizophrenia, and whether the phenomenon is an intrinsic part of the condition or a result of lifestyle and demographic factors that accompany it.

CAUSE OR EFFECT

If Jeste’s study supports the accelerated-ageing hypothesis, the next step will be to investigate how and why it occurs. Schizophrenia itself might cause accelerated ageing, but maybe some other factor, such as oxidative stress or inflammation, predisposes individuals to both conditions. Moreover, Kirkpatrick says, environmental stresses and exposures *in utero* may predispose people to both schizophrenia and accelerated ageing. The explanations are not necessarily mutually exclusive, and it’s not yet clear whether a single biological process is responsible. “I think several different mechanisms are likely to be involved,” Jeste says.

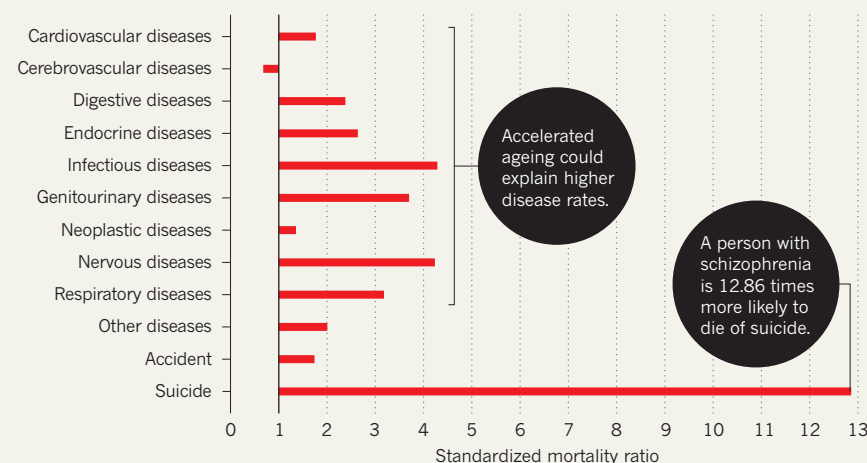
Identifying the physiological mechanisms that contribute to both ageing and schizophrenia could suggest new drug targets for schizophrenia. Confirming that oxidative stress is a common factor, for instance, could mean that antioxidant medications could help alleviate the symptoms of schizophrenia. “It will also be interesting to look at other disease models that have accelerated ageing, or are at least ageing-related, and seeing whether some of the treatments in these could be applied in schizophrenia,” says Bowie.

What started as a curious observation now has the potential to add to the understanding of a complex chronic disease and could help doctors improve the quality of life for people with schizophrenia. “These folks are dying young — that’s important,” Kirkpatrick says. “If we understood it better, perhaps we could help them.” ■

Emily Anthes is a freelance science writer based in Brooklyn, New York.

SCHIZOPHRENIA’S HEALTH BURDEN

People with schizophrenia are much more likely to die of suicide than those in the general population — but they also have higher death rates from a wide range of diseases.



1. Spelman, L. M. *et al.* *Diabetes Med.* **24**, 481–485 (2007).
2. Fernandez-Egea, E. *et al.* *Schizophr. Bull.* **35**, 437–442 (2009).
3. Fernandez-Egea, E. *et al.* *Psychosom. Med.* **73**, 643–647 (2011).
4. Nieratschker, V. *et al.* *Schizophr. Res.* **149**, 116–120 (2013).
5. Kirkpatrick, B. *et al.* *Schizophr. Bull.* **34**, 1024–1032 (2008).
6. Kochunov, P. *et al.* *Biol. Psychiatry* **73**, 482–491 (2013).

PERSPECTIVE



Retreat from the radical

Failures in the development of schizophrenia treatments don't justify the dramatic overhaul now being proposed, says **Stephen R. Marder**.

The past two decades have been a frustrating time for researchers developing drugs to treat schizophrenia and related psychotic illnesses. Ten antipsychotic medications have been introduced since the early 1990s, but large studies and meta-analyses¹ have shown that nearly all these so-called second- and third-generation antipsychotics are no more effective than agents that are more than 60 years old (the lone exception is clozapine). In the wake of these disappointing results, many companies have abandoned drug development for psychosis and other disorders of the central nervous system.

To get the field back on track, the director of the US National Institute of Mental Health (NIMH) suggested that researchers switch their efforts away from the short-term development of treatments that are only incremental improvements, and focus on understanding the core mechanisms of the disease². This approach aims to foster discoveries in genetics and the functioning of neural circuits, and proponents argue that these firm foundations are required if future treatments are to be transformative.

The NIMH is asking scientists to adopt new approaches to better define schizophrenia. The latest edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5) defines disorders by clinically observable phenomena, such as hallucinations, delusions and behavioural changes. But several illnesses can cause these symptoms, raising the risk of misdiagnosis.

The goal of the NIMH is to develop a set of diagnostic criteria based on measurable alterations in brain function. It is therefore discouraging grant applicants that use DSM-5 to categorize subjects in clinical trials. Instead, it wants investigators to use its own proposed classification system, the Research Domain Criteria (RDoC), which includes constructs that are thought to reflect basic dimensions of behaviour. Proposed RDoC constructs include negative valence systems, such as anger and fear; positive valence systems, such as reward learning; and cognitive systems, such as attention and working memory. In addition, the NIMH has shifted its focus from large clinical trials of promising interventions to smaller studies that evaluate whether drugs or other interventions engage a neuroreceptor or alter a biological process relevant to a disorder.

These are major changes in the way schizophrenia is studied, and I think the NIMH may be moving too quickly. The path from genetics to diagnosis and treatment is uncertain. Although genome-wide association studies have linked many gene variants with schizophrenia (see 'Unravelling complexity', page S6), each variant is associated with only a small effect. Thus genetics research may yield little of benefit to schizophrenia patients for decades.

Similarly, an abrupt move to a new diagnostic system is no guarantee of better results. The RDoC's constructs may help to illuminate the relationships between the functioning of neural networks and psychopathology, but they have not yet been conclusively correlated to schizophrenia symptoms. The requirement that a drug engage a neural target may be misguided; although there are a number of promising

biomarkers from neuroimaging, these have seldom been linked to the core symptoms of schizophrenia.

There is strong evidence that we are poised for substantial advances using the standard modes of research and patient categorization. Burdening investigators with an unproven diagnostic system, or a need to demonstrate the engagement of neural targets, could make this work difficult. But researchers who stick to their traditional methods may find it hard to obtain funding given the new NIMH criteria.

I am not arguing that all change is a mistake. Research has already benefited from approaches that challenge the DSM-5 categories, but without abandoning altogether the link to symptoms. For example, focusing attention on domains other than psychosis — including negative symptoms, cognition and social cognition — is likely to lead to treatments that can improve functioning and quality of life. These advances will

not prevent or cure schizophrenia, but they may mitigate the loss of function or disability associated with the illness. There are numerous active trials for drugs, devices and training interventions, so the fruits of this research are likely to reach patients in the next five to ten years.

Some of the most promising research might not fit in the new NIMH research framework. For example, there is evidence that training interventions for schizophrenia, including cognitive remediation and cognitive behavioural training therapy, can improve outcomes and enhance the functioning of brain networks³. Other evidence suggests that training interventions delivered to recently diagnosed schizo-

phrenia patients can protect the brain against the loss of grey matter in the early years of the illness⁴. These may be more than just symptomatic treatments; they could change the course of the disease⁵. More research is starting to show that neuroprotective agents might be able to alter the trajectory of the illness⁶. None of this work fits comfortably in an RDoC framework, so it might not be supported in the current environment.

The approaches endorsed by the NIMH may be able to transform the treatment and management of schizophrenia. If they pay off, the benefits will probably improve people's lives in the next few decades. But we are on the verge of less dramatic but nonetheless important improvements in the treatment of a devastating illness. Premature and radical changes in research methods may imperil these advances. ■

Stephen R. Marder is a psychiatrist at the Semel Institute for Neuroscience at the University of California, Los Angeles, and the VA Desert Pacific Mental Illness Research, Education and Clinical Center. e-mail: smarder@mednet.ucla.edu

1. Leucht, S. et al. *Lancet* **382**, 951–962 (2013).
2. Morris, S. E. & Insel, T. R. *Schizophr. Res.* **127**, 1–2 (2011).
3. Penadés, R. et al. *Biol. Psychiatry* **73**, 1015–1023 (2013).
4. Eack, S. M. et al. *Arch. Gen. Psychiatry* **67**, 674–682 (2010).
5. Stafford, M. R. et al. *BMJ* **346**, f185 (2013).
6. Beshpalov, A. et al. *Handb. Exp. Pharmacol.* **213**, 419–442 (2012).

PERSPECTIVE



Revealing molecular secrets

The more we study the genetics of schizophrenia, says **Steven E. Hyman**, the more daunting — and exciting — are the challenges we see ahead.

The genetic analysis of schizophrenia and other psychiatric disorders, such as bipolar disorder and autism, has begun to yield replicable and informative results at the molecular level^{1–3}. These successes are creating epochal opportunities to understand disease mechanisms and allow us to envisage an era of therapeutics directed at pathogenic processes — unlike current antipsychotic and antidepressant drugs that merely control symptoms.

But the emerging picture reveals great challenges ahead for translational neurobiology. Schizophrenia is highly heritable, but unlike diseases caused by a single, highly penetrant mutation (causing a phenotype with high probability), in schizophrenia the large aggregate risk attributable to genes is seen to have a polygenic basis. Disease liability is being mapped to hundreds, perhaps ultimately more than a thousand, genetic loci, each contributing a small increment of risk^{1,2,4}. Further, many of the risk-associated genetic variants (risk alleles) for schizophrenia also contribute to the risk of bipolar disorder, autism, depression and other psychiatric disorders⁵.

Copy number variants (CNVs), duplications or deletions of DNA segments affecting up to 40 genes have been found in the genomes of 2.5% of people with schizophrenia (but only 0.9% of healthy subjects). Some CNVs are potent risk factors for schizophrenia (increasing risk 10–25-fold), but even the highest-risk CNVs tell a muddled story, as they are also associated with intellectual disability and autism spectrum disorders⁶. So risk of schizophrenia is polygenic, heterogeneous across individuals, and overlaps with risk of other disorders depending on the genetic background and exposure to environmental risk factors.

As the genetic analysis of schizophrenia proceeds with ever more subjects, the list of schizophrenia-associated genes, already large, will grow further and the results will gain in certainty. However, it would be a Pyrrhic victory if the resulting list of genes yielded no biological insights that ultimately led to effective treatments.

It is not yet clear how to achieve the desired outcomes. Current tools to study gene function have been optimized to investigate highly penetrant mutations with large effects on phenotype. Typically, researchers study the effects on phenotype that result from either knocking a gene out entirely or inserting a highly penetrant allele into an animal model. But given the low penetrance of schizophrenia-associated alleles, and their ability to contribute to different diseases depending on the genetic background of the organism, inserting one or even several into an animal model may yield a phenotype that is ambiguous with respect to human disease — or no phenotype at all.

The large number of genes involved in schizophrenia, and their combinatorial interactions, require models that permit efficient testing of many different hypotheses. Moreover, many loci that contribute to polygenic disorders including schizophrenia occur in non-coding parts of the genome, which are less well conserved in evolution than segments of DNA that encode proteins. For this reason, and because

many features of schizophrenia pathogenesis occur in evolutionarily recent neural cell types and circuits, it is critical to take evolution into account in matching a model to the question being asked. Given such constraints, the kinds of genetic mouse models that have been used for decades to study diseases caused by single, highly penetrant mutations are likely to have severe limitations for the study of schizophrenia.

According to a model for how polygenic risk factors operate, the products of allelic variants appear in key protein networks and biochemical pathways in cells and perturb their function, causing disease. Each individual is likely to have a different combination of variants that produce pathological tweaks in these networks. The ability to study these networks is more likely than studying individual genes to yield insight into schizophrenia mechanisms and to suggest drug targets. For example, a possible protein network arising from the schizophrenia genetic data is involved in the postsynaptic specialization of excitatory neurons^{2–4}.

Despite a wealth of genetic findings, further investment in gene discovery is needed. Each risk allele discovered identifies a gene with some role in pathogenesis, and provides a new piece in the jigsaw puzzle from which important networks and pathways emerge⁴.

A more difficult question concerns the selection of living systems, cells or animal models for the functional analysis of alleles and pathways. Thanks to advances in genome engineering, it should be possible to insert DNA sequences of interest into many models, and enable researchers to engineer many mutations at once⁷. Models may include induced pluripotent cell lines and human

embryonic stem-cell lines reprogrammed into neurons to study molecular and cellular mechanisms, and serve as a platform for chemical screens. Progress is being made in reprogramming such cells into relevant neuronal and glial cell types, although questions remain over the cell types most involved in various psychiatric disorders.

Invertebrates, zebrafish, mice and non-human primates may all prove to be useful genetic models. It is too early to say what combination of approaches will work. All we know is that the emerging genetic discoveries will call for innovation and a willingness by foundations, government funding agencies and scientists to try new things. ■

Steven E. Hyman is director of the Stanley Center for Psychiatric Research at the Broad Institute of Harvard and Massachusetts Institute of Technology in Cambridge, Massachusetts. He is a former director of the US National Institute of Mental Health. e-mail: seh@harvard.edu

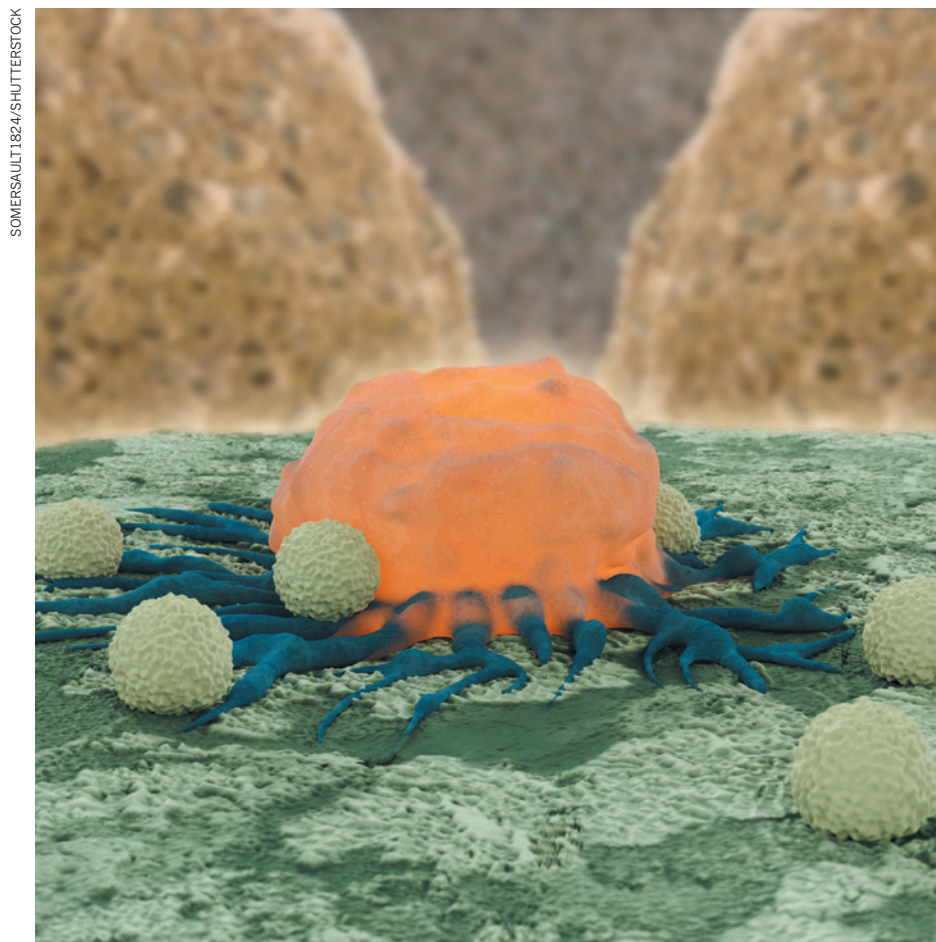
1. Ripke, S. *et al.* *Nature Genet.* **45**, 1150–1159 (2013).
2. Purcell, S. M. *et al.* *Nature* **506**, 185–190 (2014).
3. Fromer, M. *et al.* *Nature* **506**, 179–184 (2014).
4. McCarroll, S. A. & Hyman, S. E. *Neuron* **80**, 578–587 (2013).
5. Cross-Disorder Group of the Psychiatric Genomics Consortium. *Nature Genet.* **45**, 984–994 (2013).
6. Kirov, G. *et al.* *Biol. Psychiatry* **75**, 378–385 (2014).
7. Cong, L. *et al.* *Science* **339**, 819–823 (2013).

CAREERS

TURNING POINT Solar System findings propel geochemist back to China **p.141**

EQUALITY More women sought to meet UK tech-sector needs **p.141**

NATUREJOBS For the latest career listings and advice www.naturejobs.com



White blood cells (grey) can be engineered to attack specific types of cancer (red).

TRANSLATIONAL RESEARCH

Cancer killers

Promising results in cancer immunotherapy offer growing opportunities – and challenges – in translational research.

BY RACHEL BERNSTEIN

Michel Sadelain recalls his colleagues' response to his postdoctoral research project: a combination of amusement and incredulousness. He wanted to eschew the vaccine-based approaches that were common at the time, and instead kill cancer cells by tinkering with the immune system genetically.

Twenty years later, his insight seems to have

been vindicated. Early successes in treating people with cancer using engineered immune cells called T cells have breathed new life into the once-maligned field of cancer immunotherapy (see page 24), and the result has been an influx of funding, as well as growing opportunities in academia and industry. Indeed, Sadelain's gamble has brought him career success: he is now director of the Center for Cell Engineering and Gene Transfer at Memorial Sloan

Kettering Cancer Center in New York City.

Until now, the field has "always been a little bit of a stepchild", says cancer immunotherapy pioneer Steven Rosenberg, head of the tumour immunology section at the US National Cancer Institute in Bethesda, Maryland. Researchers have been reluctant to get involved because it is difficult to genetically engineer mammalian cells and because cancer cells are very similar to normal cells, making them particularly difficult to target with immunotherapies.

But during the past few years, promising results from early clinical trials have shown that therapies based on T-cell engineering can force some stubborn cancers into remission. With these successes, "cancer immunology has come to the forefront both in immunology and in cancer treatment", Rosenberg says.

Numbers documenting the rise in prominence are hard to come by. But an auspicious sign is that the amount of money awarded to the field by the American Association for Cancer Research in Philadelphia, Pennsylvania, has risen from just US\$400,000 in 2011 to \$27.2 million in 2013. For the foreseeable future, the field is rife with opportunity.

RESEARCH RESURGENCE

The results of these early clinical trials — done mostly in blood cancers — are not just promising for patients, they are also opening up research avenues in the laboratory. "What we've done now is established that this is a valid new field of cancer therapy," says Marco Davila, who in January took up a position studying the potentially toxic side effects of T cell-based therapies at Vanderbilt University Medical Center in Nashville, Tennessee. "This is going to lead to an entire new set of questions that need to be evaluated in a very basic manner, as well as in a translational manner."

A key focus is on expanding the success of immunotherapies. The trials that have been successful so far have mainly targeted B cells, which are white blood cells that become mutated in B-cell leukaemia and lymphoma. T cells can be engineered to kill cells that display the CD19 protein, which is found on both healthy and cancerous B cells. This non-specific approach is acceptable because B cells are not crucial to survival, but investigators are working to identify targets for other cancers and to develop approaches that target just cancerous cells. They are also trying to work out why some patients do not respond to B-cell immunotherapy.

The fledgling immunotherapy field is being led by a handful of key players, including ►

► Rosenberg, Carl June of the University of Pennsylvania in Philadelphia and Laurence Cooper of the University of Texas MD Anderson Cancer Center in Houston. But young researchers such as Davila are joining the effort — and with diverse questions still to be answered, scientists from a variety of backgrounds are in demand.

MULTIPLE ENTRY POINTS

“There are so many different ways to get [into the field] — you don’t have to limit yourself to doing cancer immunotherapy during graduate school or your postdoc,” says Marcela Maus, who works with June and is the director of medical affairs for translational research at Penn’s Abramson Cancer Center. She cites entry points such as virology, T-cell biology and cancer signalling. Maus herself did not start in the field until she was a graduate student, having conducted evolutionary-biology research on *Escherichia coli* as an undergraduate. She found that bacterial concepts such as selection pressure apply to cancer as well — and to her current work on interactions between T-cell therapies and other cancer treatments.

One of June’s postdocs, Joseph Fraietta, switched from his graduate work on HIV after reading about the exciting progress in cancer immunotherapy. And Rosenberg says that there is a growing demand for geneticists and bioinformaticians who can analyse tumour-genome data to identify potential immune targets. “There’s a lot known about immunology, and a lot known about genomics, but the two have not been tied together in a meaningful, applicable fashion until very recently,” he says.

Those interested in developing cancer immunotherapies should carefully consider the clinical facets of the work. “To apply research to patient-related problems, you either need training in medicine or very close collaborations with people who are trained in medicine,” Rosenberg says. Ton Schumacher, a group leader at the Netherlands Cancer Institute in Amsterdam, recommends acquiring a good understanding of immune reactions and knowledge of how to translate concepts into clinical trials. Many in the field earn a joint MD–PhD degree, but an MD is not required as long as researchers interact with clinicians to make sure that their work is relevant to patients.

One reason that cancer immunotherapies are so promising is that they are engineered for individual patients, using a person’s own T cells to specifically target the cancer. But if cancer

immunotherapy is to succeed, practitioners will have to find ways to scale up manufacturing capabilities despite their personalized nature. “This is not a well-known path to making drugs,” says Sadelain. “There are many components of manufacturing and distribution for which there is no precedent.”

To help to work out such issues, pharmaceutical and biotechnology companies are seeking academic partners, which could pave the way for opportunities for early-career scientists. For example, June teamed up with Novartis a few years ago after successfully testing an immune therapy that used genetically modified T cells called chimaeric antigen receptors (CARs). Since then, Novartis and Penn have jointly established the Center for Advanced Cellular Therapies on the university’s campus to further develop CAR technologies, and Novartis is hiring cancer-immunotherapy researchers.

And in March last year, researchers at the Center for Cell and Gene Therapy at Baylor College of Medicine in Houston, Texas, a hub for cancer-immunotherapy work, teamed up with two biotech companies, Celgene in Summit, New Jersey, and bluebirdbio in Cambridge, Massachusetts. And in December, Sadelain worked with scientists at the Memorial Sloan Kettering Cancer Center and two other cancer-research institutions to form Juno Therapeutics in Seattle, Washington, which will work closely with academic researchers to develop T-cell therapies. All this means that postdocs now have opportunities beyond just university-based research, June notes.

Lenka Hurton, a graduate student who is working on improving the persistence of the engineered T cells in Cooper’s laboratory, says that she hopes to pursue an industry job after completing a postdoc and is encouraged that companies such as Novartis are recruiting. “I’m anticipating that there will be other companies that do the same, assuming that the clinical trials continue to show promising results,” she says.

For those working at a university, collaborations with industry can provide valuable exposure and resources. “I’m really getting a lot of bang for my buck because I’m in a semi-academic, semi-industrial postdoc,” says Fraietta, who is involved in the Penn–Novartis work. “I have this hybrid situation where I get to see the best of both worlds.”

TRANSLATIONAL CHALLENGES

Many in the field agree that those who stay at academic institutions must learn how to conduct translational research in a culture and environment that may be unaccustomed to supporting such work. There are, for example, practical considerations. Clinical trials generally require a longer time frame than do conventional grant cycles — it can take two to three years just to get a trial off the ground. “If you need to write grants every year, that’s going to be tough,” Maus says. And although trials that



Michel Sadelain is an immunotherapy trail-blazer.

result in a blockbuster drug can be great career boosters, they constitute only a tiny part of the research needed to develop effective therapies. Although essential for safely manufacturing new therapies, projects that focus on the painstaking trouble-shooting, optimization and process-development work do not usually garner high-impact publications, which may discourage ambitious young researchers.

Davila, a new professor who is establishing his own research group, has a strategy to stay on track. “I’m sticking with very finite goals that are going to be meaningful, that are going to contribute to the field,” he says. He chose his project with care: developing an animal model of cytokine release syndrome, a potentially fatal side effect of CAR T-cell treatment. He hopes that this approach will enable him to complete projects within the short time frame required for publishing and obtaining grant funding in an academic setting.

First and foremost, immunotherapy researchers focus on creating and improving effective treatments — not just pursuing intriguing phenomena. Hence, they might not have the luxury of dallying in research just to sate their intellectual interest. The aim is always to push forward towards better treatments.

Cooper therefore encourages graduate students to earn their PhD by investigating the more fundamental research questions in cancer immunotherapy, then waiting until their postdoc or later to pursue the translational angle. Maus followed this track. “As a graduate student, I wasn’t really able to do the translational aspects directly,” she recalls. “It was frustrating at the time — but now I realize it was really important to focus on basic science.” So far, Maus and others who have been able to balance these priorities are reaping the career rewards. “For young people getting into the field,” says Cooper, “there’s never been as good a time as this.” ■

Rachel Bernstein is a science writer based in San Francisco, California.



“I have this hybrid situation where I get to see the best of both worlds.”

Joseph Fraietta

TURNING POINT

Liping Qin

In June, Liping Qin will become the first person from China to be presented with the European Association of Geochemistry's Houtermans Award, which recognizes a single exceptional contribution to geochemistry by an early-career scientist.

Why did you move to the United States?

I knew I wanted to be a scientist — I was always experimenting at home with chemistry or physics sets. I was accepted at the University of Science and Technology of China (USTC) and went to the geochemistry department. But I looked for graduate opportunities in the United States because geochemistry in China in the late 1990s was not that robust — it didn't have a big presence in the field then.

What led you to cosmic chemistry?

I decided to come to the University of Chicago in Illinois to join a project on finding geochemical markers, such as ratios of iron to manganese, that could help us to trace magma back to Earth's mantle. I also attended seminars about the cosmos and stars, and started talking to cosmic chemist Nicolas Dauphas, which led to a project using radioactive isotopes to date meteorites. I worked on a special group of meteorites made up mostly of iron, which are thought to be remnants of the cores of planetesimals, small bodies that lead to the formation of planets.

How was this a turning point in your career?

I found that the first-generation planetesimals formed within a few million years of the Solar System's birth. Before this discovery, researchers believed that these planetesimals formed later, as the Solar System developed, but our work helped to confirm the earlier timing. During that project, I found isotopic anomalies of tungsten in meteorites. I got very interested in using isotopes to understand what the stellar environment was like in the early Solar System.

Where did you set your sights for a postdoc?

My first priority was to get to the Carnegie Institution for Science in Washington DC because it gives fellows the freedom to explore any research that interests them. While I was there, I continued to work on isotopic anomalies of other elements, such as chromium and barium, to trace the early stellar environment. Working with so many great scientists helped me to understand the



big questions in this field, including the age and formation of the Moon, Earth's building blocks and the timing and mechanism of water's appearance on Earth.

Where did you go next?

I chose to go to Lawrence Berkeley National Laboratory (LBNL) in California and do biogeochemistry work, studying Earth-surface processes, because I wanted to challenge myself. LBNL wanted me to apply my isotope skills to biological processes — specifically, the bioremediation of hexavalent chromium, which is a very toxic carcinogen that has contaminated aquifers. I expanded my area of expertise and my circle of collaborators. It was a good move.

Were you eager to return to China?

I knew I would go back one day, and in 2012, I had the opportunity to do so — I received a generous funding package to start my own lab at the USTC. I've continued to work on cosmo-chemistry here because it's what interests me and I'm good at it. I'm also using the same skills on some biological and environmental questions, including the use of metal isotopes to trace enzyme processes associated with metal movement.

How will the award help you professionally?

I hope that it will help to grow our research group so that my students have more opportunities and resources. I have five graduate students now. It was the right time for me to return to China to do planetary chemistry, because there is so much student interest in connection with China's plans to launch a lunar-sampling mission. My students want to study cool rocks. ■

INTERVIEW BY VIRGINIA GEWIN

EQUALITY

Mind the gender gap

A report by a UK non-profit agency says that UK businesses need more technicians, manufacturing leaders and engineers, and that more women must earn degrees in physical sciences to meet the demand. *Target 2030*, from the National Centre for Universities and Business in London, notes that just 9.5% of science and engineering professionals in the United Kingdom are women. In 2013, women represented about one-fifth of physics students at pre-university-qualification levels, and the gender gap in the subject is widening. The report notes that women with physical-sciences degrees have higher employability rates than their male counterparts, and calls for a campaign to attract more female students into the higher-education pipeline in physics, maths and related disciplines.

VISAS

Stay for a while

Some legislators in Sweden want to adopt a proposal by 1 June that would allow foreign doctoral students to remain in the country temporarily to work or to launch a company after earning their degree. Furthermore, PhD students who have studied in Sweden for at least four of the past seven years would qualify for permanent residence. Sweden's parliament has yet to vote on the legislation. Meanwhile, the European Commission and other European Union (EU) legislative bodies are considering a proposal that would allow doctoral students and researchers in non-native countries to stay in EU member states after completing their degree or research contract. Legislative action could come this autumn.

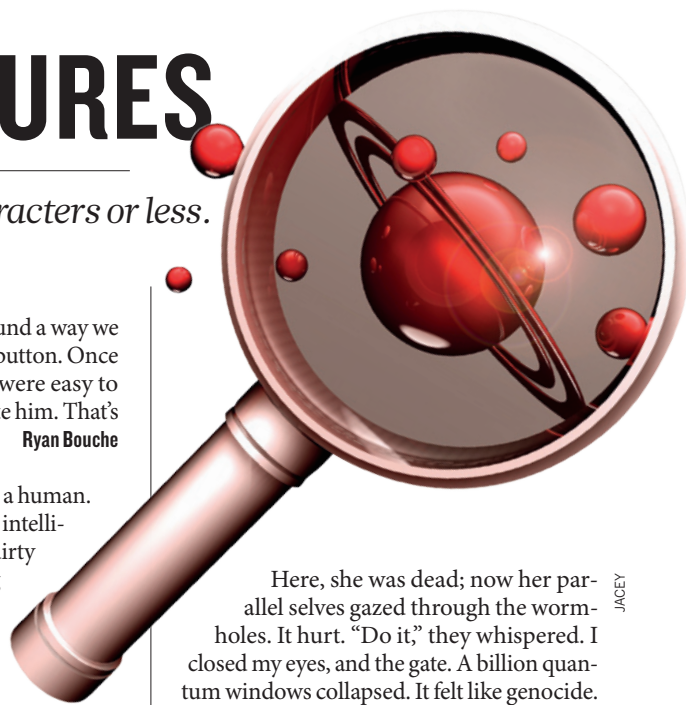
FUNDING

UK postgrad bonanza

The UK government is committing £496 million (US\$820 million) towards 111 postgraduate training centres in physical sciences, maths and engineering to plug skills gaps in industry. UK chancellor George Osborne said on 19 March that the government will spend £106 million over five years on 20 new centres, to top up its earlier commitment of £390 million towards 91 centres. Industry has provided another £374 million. The Engineering and Physical Sciences Research Council is administering the funds mainly as graduate studentships.

microFUTURES

Science fiction in 200 characters or less.



JACEY

At the start of this year, Futures ran a competition challenging readers to write a story in just 200 characters. After a difficult judging session, we're pleased to award first prize to Catherine Rastovski, who wins a year's subscription to *Nature* plus a gift voucher for the *Futures 1* eBook. Five runners up also receive gift vouchers for *Futures 1*. Here are the winning entries, plus a selection of stories that impressed the judges.

FIRST PRIZE

I pass your empty chair every day. Across the room sits the computer, your voice, your face locked inside. I ache to bring you to life, but fear keeps me in my chair. What if you say no?

Catherine Rastovski

RUNNERS UP

Cynthia was learning faster than anyone had predicted. The apparently totipotent cells continued to proliferate at an exponential rate. Today would be the last time they would refer to her as a mouse.

Adam Flanders

The champagne tasted unpredictably sour, the time traveller thought, as the press cameras flashed. Even the world-record lottery win was nowhere near enough to repair the broken machine.

Arran Froid

off that switch, Professor! Your time machine can't travel back in time past the moment of its own creation and instead will trap the Universe in an endlessly recursive time-like loop! Take your hands

Judith Reeves-Stevens

Silence. I floated around through corridors, filled with debris. No flickering lights, no air, no sound, just memories I could not grasp. Alone. I released my air supply and became part of the debris.

Anssi Sajama

The robot knows when you turn it off. Those last picoseconds, as the current drains from its circuits, are a living Hell. When you turn it back on, it will never be the same. But it will remember ...

Chaim Schramm

COMMENDED

It took a while, but we finally found a way we were different. He had no belly button. Once we had one difference, others were easy to find. That's when I started to hate him. That's why he had to go.

Ryan Bouche

I forgot what it was like to touch a human. Less beautiful, less obliging, less intelligent than my companions — a dirty mess of bacteria. Discovering his imperfections. I resolved to never try again.

Adam Dunn

I was the first to make contact, but they spoke not with their mouths, only with their eyes.

And their eyes said: "Fear us, for we are your demise."

Cody Frederick

As robots perform the laboratory work, it is cheaper just to electronically stimulate areas of the student's brain associated with frustration and failure and then, after 3 years, call them doctor.

Gavin Garland

The Prof smiled. "I've isolated the plague vector! Now we can kill it." I sighed with relief, "I'll tell the others." He held my arm as I turned away. "Hang on, old chap. Amy, pass me the syringe."

John Gilbey

I managed to find the most-wanted gift of the year. Finally, I found a way to surprise him. The real surprise came when it unwrapped itself and knocked loudly on the bedroom door early this morning.

Jamie Hadac

For sale: one cat-litter box, never used; or not. Contact E. Schrödinger.

Josés Ho

Sorry—I overslept/sleep tomorrow—missed invite 2 come 2 yr quantum state. Let's try again—Googlemas 2 years ago. I'll bring/brought nano-geese + 2096! intox/ferment. C/saw U 2 years ago. luv-u ^

Ken Krug

We gather, receiving the voice of a deserted explorer. "I have found the lunar colony," she says across the void. "Their logs simply state, 'We die alone.'" Eyes downcast, we hear the broadcast end.

Alasdair MacLeod

Despite her growing love for him the mission demanded it happen now. He sat on the bed and placed his hand on her leg, she turned inside out enveloping the president. Endocytosis won the ten-day war.

Dan Nicholls

Here, she was dead; now her parallel selves gazed through the wormholes. It hurt. "Do it," they whispered. I closed my eyes, and the gate. A billion quantum windows collapsed. It felt like genocide.

Aude Oliva

The city is microbial more than it is concrete. Licking the wall is today's graffiti and in saliva swirls the future of urban planning. She spat into the web, and smiled. The diagram shifted slightly.

Poeia Onomato

Still orbiting Earth, a century later: a little girl's shoes.

Charles Pellegrino

"Rest assured that you'll eventually be vindicated," said the time traveller, handing the medal to a bewildered Turing. "Regrettably, beyond giving you this, we are not allowed to alter the past."

Gary Shockley

The Sentence It was a novel trial, the first of its kind, the charge was illegal human cloning, creating a sentient human in laboratory, the verdict was guilty, ironically the sentence was: life.

Nathan Susnik

Years of cheap computer time allowed the TKLabs GeneCoin project to decode all the 'junk' DNA in the human genome. Now a lone tech sees the output flashing on the screen: "Intentionally left blank" ...

The Moletrap Collective

I lost my arms at work. New arms made from tissue scaffolds take too long to grow. No work, no money. So, my boss fused a used pair of robotic arms onto me. My robot overlords finally promoted me!

Mark Teo

Thorus remained deathly still as he frantically searched his pocket for the laser gun. He hadn't seen the ship's cat for a few days now, and that thing he'd just seen had huge teeth ...

Mick Watson

➔ **NATURE.COM**

Follow Futures:

🐦 @NatureFutures

📘 go.nature.com/mtoodm